

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|--|-----------|---|
| (51) International Patent Classification ⁶ : C12Q 1/68 | A1 | (11) International Publication Number: WO 98/14616 (43) International Publication Date: 9 April 1998 (09.04.98) |
| (21) International Application Number: PCT/US97/17877 (22) International Filing Date: 3 October 1997 (03.10.97) (30) Priority Data: 08/726,090 4 October 1996 (04.10.96) US (71) Applicant: PERSEPTIVE BIOSYSTEMS, INC. [US/US]; 500 Old Connecticut Path, Framingham, MA 01701 (US). (72) Inventors: HAFF, Lawrence, A.; 167 Deer Run Road, Wilton, CT 06897 (US). SMIRNOV, Igor Pavlovich; 50 Davis Avenue, Brookline, MA 01760 (US). (74) Agent: TURANO, Thomas, A.; Testa, Hurwitz & Thibault, LLP, High Street Tower, 125 High Street, Boston, MA 02110 (US). | | (81) Designated States: European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> |
| (54) Title: METHODS FOR DETERMINING SEQUENCE INFORMATION IN POLYNUCLEOTIDES USING MASS SPECTROMETRY (57) Abstract The invention relates to method for determining sequence information in polynucleotides by combining the recent disparate technologies of mass spectrometry and polynucleotide hybridization, amplification, extension and/or ligation techniques. Broadly, in a first step, the method for determining sequence information in a sample polynucleotide includes hybridizing with a sample nucleotide one or a mixture of oligonucleotide probes having a nucleotide sequence complementary to a portion of the sample polynucleotide, thereby forming a complex. Then, in a second step, the complex is contacted with at least a member selected from the group consisting of nucleosides, dideoxynucleosides, polymerases, nucleases, transcriptases, ligases and restriction enzymes to alter at least a subset of said oligonucleotide probes. In a third step, the method provides for determining the molecular weight of at least the subset of altered probes by mass spectrometry and inferring the sequence information of the sample polynucleotide therefrom. | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakhstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LJ | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

METHODS FOR DETERMINING SEQUENCE INFORMATION IN POLYNUCLEOTIDES USING MASS SPECTROMETRY

Field of the Invention

The invention generally relates to methods for the determination of sequence information in polynucleotides and more specifically for the identification of nucleotide polymorphisms using mass spectrometry.

Background of the Invention

Gene probe assays which depend upon the binding of DNA probes with their complementary base pairs on a target molecule are among the most common assays used by molecular biologists. Such assays require a great deal of time to bind the probe to a target DNA, to remove any excess unhybridized probe and to analyze the hybridized probe.

The time required to perform these assays and the difficulties associated in performing these assays make such assays inefficient in searching for single nucleotide polymorphisms (SNPs). An SNP is a change (deletion, insertion or substitution) in any single nucleotide base in the region of the genome of interest. Because SNPs occur so frequently in the human genome (about once in every 500 bases), SNPs are useful markers in studying the human genome.

The detection of SNPs is typically performed using automated DNA sequencers. However such sequencers are generally not well utilized in determining a single base change within a 500 base sequence. As a result an SNP search using a sequencer is slow and expensive.

The recent development of Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS) has permitted small amounts (about 5 femtomoles) of DNA to be analyzed to extreme resolution (one dalton accuracy). The sensitivity and mass resolution of MALDI-TOF mass spectrometry fall off with increasing mass, so that the current upper practical limit for analysis of DNA is between about 50 to 100 nucleotide bases. Samples of 500 nucleotide bases have been analyzed with MALDI-TOF but with poor resolution and sensitivity.

The present invention relates to a method for quickly determining polynucleotide sequences, with low labor intensity and a low cost per SNP assay.

- 2 -

Summary of the Invention

In its broadest aspect, the invention provides a family of techniques for determining sequence information in polynucleotides by combining the recent disparate technologies of mass spectrometry and polynucleotide hybridization, amplification, extension and/or ligation techniques. Broadly, in a first step, the method for determining sequence information in a sample polynucleotide includes hybridizing with a sample polynucleotide one or a mixture of oligonucleotide probes having a nucleotide sequence complementary to a portion of the sample polynucleotide, thereby forming a complex. Then, in a second step, the complex is contacted with at least a member selected from the group consisting of nucleosides, dideoxynucleosides, polymerases, nucleases, transcriptases, ligases and restriction enzymes to alter at least a subset of said oligonucleotide probes. In a third step, the method provides for determining the molecular weight of at least the subset of altered probes by mass spectrometry and inferring the sequence information of the sample polynucleotide therefrom.

In one embodiment the invention relates to a method for identifying a putative point mutation in a polynucleotide using mass spectrometry. The method includes hybridizing a primer to the polynucleotide having a putative point mutation such that the 3'- end of the primer is hybridized adjacent the putative point mutation; extending the 3'- end of the primer by adding a single nucleotide base matching the putative point mutation of the polynucleotide, thereby forming an extended primer; and identifying the added nucleotide base on the extended primer using mass spectrometry. In one embodiment the mass spectrometry used is matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. In one embodiment the step of hybridizing the primer includes the steps of providing a primer having a sequence that hybridizes to a polynucleotide sequence next to a putative point mutation; heating the polynucleotide to a temperature sufficient to melt the polynucleotide; and cooling the polynucleotide to a temperature sufficient to hybridize the primer to the polynucleotide, thereby forming a duplex wherein the 3'- end of said primer is adjacent the putative point mutation. In one embodiment the step of extending includes providing at least one dideoxynucleoside triphosphate base (ddNTP) and providing a polymerase. In one embodiment, the polynucleotide is a DNA and the polymerase is a DNA polymerase. In another embodiment, the polynucleotide is an RNA and the polymerase is a reverse transcriptase.

Another embodiment of the invention exploits the use of a primer having a mass tag on the 5'- end of the primer. In one embodiment the mass tag is a thymidine DNA or polythymidylate segment, T_n , wherein n is an integer ranging from 1 to 40.

In yet another embodiment the invention relates to a multiplex method for identifying multiple putative point mutations in a polynucleotide using mass spectrometry. The method, in one embodiment,

- 3 -

includes hybridizing a multiplicity of primers to a polynucleotide having multiple putative point mutations, the 3'- end of each primers being hybridized adjacent each putative point mutation; extending the 3'- end of each primers by adding a single nucleotide base to each primer, each single nucleotide base matching each putative point mutation of the polynucleotide, thereby forming a multiplicity of extended primers; and
5 identifying the added nucleotide base on each extended primers using mass spectrometry. In one embodiment, the polynucleotide is a DNA and the polymerase is a DNA polymerase. In another embodiment, the polynucleotide is an RNA and the polymerase is a reverse transcriptase.

Another embodiment of the invention includes a method for identifying a putative point mutation in a DNA using mass spectrometry, including the step of hybridizing two oligonucleotide probes to a single
10 stranded DNA having a putative point mutation such that the two probes together overlap the putative point mutation, ligating the two probes into a ligated single strand oligonucleotide only when the probe overlapping the putative point mutation has the nucleotide base complementary to the nucleotide base of the DNA at the putative point mutation and determining the identity of the nucleotide base matching the putative point mutation on the ligated oligonucleotide using mass spectrometry. In one embodiment the
15 mass spectrometry is performed using matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry.

In still yet another embodiment the invention relates to a combinatorial method for determining sequence information in a region of a polynucleotide using mass spectrometry. In one embodiment the method includes, hybridizing a primer to a polynucleotide having a region of interest such that the 3'- end of
20 the primer being hybridized is adjacent the region of interest and extending the 3'- end of the primer by adding a multiplicity of nucleotide bases matching the region of interest of the polynucleotide, thereby forming an extended primer. In one embodiment the extending step includes providing a multiplicity of nucleotide bases selected from the group consisting of deoxynucleoside triphosphate (dNTP) and dideoxynucleoside triphosphate (ddNTP), provided that at least one member of ddNTP is present, and
25 providing a polymerase in a buffer solution. The identity of the added nucleotide bases on the extended primer is then determined using mass spectrometry. In one embodiment, the polynucleotide is a DNA and the polymerase is a DNA polymerase. In another embodiment, the polynucleotide is an RNA and the polymerase is a reverse transcriptase. In another embodiment the invention includes providing an exonuclease to the mixture resulting from the extension thereby producing a ladder sequence defining the
30 mixture of extended primers. In one embodiment the exonuclease is selected from the group consisting of: phosphodiesterase type I, exonuclease I, exonuclease III, exonuclease V, exonuclease VII, and DNA polymerase III.

In another embodiment the invention relates to a multiplex combinatorial method for determining sequence information in multiple regions of a polynucleotide using mass spectrometry. The method includes hybridizing a multiplicity m of primers to a polynucleotide having multiple regions of interest such that the 3'- end of each of the primers is hybridized adjacent each region of interest, and extending the 3'-
5 end of each of the primers by adding a multiplicity of nucleotide bases to each primer matching each region of interest of the polynucleotide, thereby forming a multiplicity of extended primers. In one embodiment the extension step includes providing a multiplicity of nucleotide bases selected from the group consisting of deoxynucleoside triphosphate (dNTP) and dideoxynucleoside triphosphate (ddNTP), provided that at least one member of ddNTP is present and providing a polymerase in a buffer solution. In one embodiment,
10 the polynucleotide is a DNA and the polymerase is a DNA polymerase. In another embodiment, the polynucleotide is an RNA and the polymerase is a reverse transcriptase. The identity of the added nucleotide bases on each extended primers is determined using mass spectrometry.

Brief Description of the Drawings

The foregoing and other objects, features and advantages of the present invention, as well as the invention itself, will be more fully understood from the following description of preferred embodiments, when read together with the accompanying drawings, in which:

5 Fig. 1 is a highly schematic diagram of an embodiment of the method for identifying a single putative point mutation or polymorphic site in a DNA using PCR;

 Fig. 2 depicts the results of a single nucleotide polymorphism (SNP) assay using primer extension reaction with: ddA, line A; ddC, line B; ddG, line C; ddT, line D; and the four ddNTPs, line E;

10 Fig. 3 is a highly schematic diagram of an embodiment of the assay used to determine heterozygous single nucleotide polymorphisms;

 Fig. 4 is a diagram of a mass spectrum of the results of a single nucleotide polymorphism assay performed, in the presence of all four ddNTPs, on an heterozygous sample having C and T at the variable site;

15 Fig. 5 is a highly schematic diagram of an embodiment of an assay used to determine multiple single nucleotide polymorphisms using mass tags;

 Fig. 6a is a mass spectrum of five primers having mass tags used in performing a multiplex single nucleotide polymorphism assay for BRCA1 single point mutations R355, R377, R400, R421, and R447;

 Fig. 6b is a mass spectrum of the results of the multiplex assay performed in the presence of the five primers whose mass spectrum is shown in Fig. 6a;

20 Fig. 6c is a mass spectrum of seven primers having mass tags used in performing a multiplex single nucleotide polymorphism assay for BRCA1 single point mutations F320, R355, R377, R400, R421, R480 and R447;

 Fig. 6d is a mass spectrum of the results of the multiplex assay performed in the presence of the seven primers whose mass spectrum is shown in Fig. 6c;

25 Fig. 6e is a mass spectrum of seven primers having mass tags used in performing a multiplex single nucleotide polymorphism assay for BRCA1 single point mutations R355, R377, R400, R421, and R447 and for *lacZ* single point mutation (b2-1);

- 6 -

Fig. 6f is a mass spectrum of the results of the multiplex assay performed in the presence of the seven primers whose mass spectrum is shown in Fig. 6f;

Fig. 7 is a highly schematic diagram of an embodiment of the method for identifying a single putative point mutation in a double stranded DNA using ligase;

5 Fig. 8 is a mass spectrum of the results of a ligase assay using an oligonucleotide probe with a mass tag;

Fig. 9a is a mass spectrum of the results of a ligase assay of heterozygous DNA using multiple oligonucleotide probes with mass tags;

10 Fig. 9b is a highly schematic diagram of an embodiment of the method for identifying a single putative point mutation using ligase in an heterozygous DNA sample;

Fig. 10 is a highly schematic diagram of an embodiment of a combinatorial assay used to determine the presence of more than one nucleotide mutation at a putative mutation site;

Fig. 11 is a mass spectrum of the results of the combinatorial assay using a polymerase;

15 Fig. 12 is a highly schematic diagram of a combinatorial assay used to determine the presence of more than one nucleotide mutation at a putative mutation site followed by digestion by an exonuclease;

Fig. 13 is a highly schematic diagram of a multiplex combinatorial assay on a DNA having multiple putative mutation sites of more than one nucleotide base mutation;

Fig. 14 is a highly schematic diagram of an embodiment of an instrument used to practice the invention; and

20 Fig. 15 is a flow diagram of an embodiment of an algorithm used in conjunction with the instrument shown in Fig. 14.

Detailed Description of the Invention

25 In its broadest aspect, the invention provides a family of techniques for determining sequence information in polynucleotides by combining the recent disparate technologies of mass spectrometry and polynucleotide hybridization, amplification, extension and/or ligation techniques. Broadly, in a first step, the method for determining sequence information in a sample polynucleotide includes hybridizing with a sample polynucleotide one or a mixture of oligonucleotide probes having a nucleotide sequence

- 7 -

complementary to a portion of the sample polynucleotide, thereby forming a complex. Then, in a second step, the complex is contacted with a member selected from the group consisting of nucleosides, dideoxynucleosides, polymerases, nucleases, transcriptases, ligases and restriction enzymes to alter at least a subset of said oligonucleotide probes. In a third step, the method provides for determining the molecular weight of at least the subset of altered probes by mass spectrometry and inferring the sequence information of the sample polynucleotide therefrom.

The invention also relates to methods for the identification of nucleotide polymorphisms using mass spectrometry. In the following detailed description of the invention, the terms "putative point mutation" or "single nucleotide polymorphism" are used interchangeably and are intended to mean any modification in a polynucleotide sequence which result in a single nucleotide base mutation, such as an insertion, a deletion or a substitution of a single nucleotide base. The term "putative mutation site" is intended to mean any modification in a polynucleotide sequence which results in a mutation of multiple nucleotide bases, resulting from an insertion, a deletion or a substitution of multiple nucleotide bases or combination thereof. The term "region" is intended to mean a portion of the sequence of a polynucleotide.

In brief overview and referring to Fig. 1, in one embodiment, the process of identifying a putative single point mutation 10 in a DNA 20 includes first (Step 1) separating the strands of the double stranded DNA into two single strands; one of which being a target strand 22 having the putative point mutation 10 of interest. Next (Step 2) a primer sequence 30 having a sequence of bases which is complementary to the bases on the target strand 22 up to and including the base 24 adjacent the putative point mutation 10 is annealed to the target strand 22.

Each of the possible complementary bases 36, 36', in the form of dideoxynucleotides, is added (Step 3) along with a DNA polymerase to the primer-target DNA complex. The dideoxynucleotide base 36 which is complementary to the putative point mutation 10 is incorporated into the primer sequence 30 (Step 4) to form an extended primer 30'. Because the primer 30' is now terminated with a dideoxynucleotide base 36, no further additions are possible.

The primer sequence 30' with the added dideoxynucleotide base 36 is next analyzed with a mass spectrometer (Step 5). Note that when using MALDI-TOF mass spectrometry, the primer-target DNA complex will fully dissociate under the laser desorption ionization process of the MALDI-TOF mass measurement. The mass of the primer 30' with the additional mass of the dideoxynucleotide base 36 is determined (comparative mass shown in Table 1), and, by subtracting the original mass of the primer 30, the mass of the dideoxynucleotide base 36 and, hence, the type of dideoxynucleotide base 36 are determined. By knowing the identity of the added dideoxynucleotide base 36, the identity of the

- 8 -

complementary putative point mutation 10 is therefore deduced according to the rules of Watson-Crick base pairing.

Table 1: Dideoxy and Deoxy-Nucleotide bases

| Dideoxy Base | Formulae | Mass | Deoxy Base | Formulae | Mass |
|--------------|-----------------------|---------|------------|-----------------------|---------|
| -ddC | $C_9H_{12}N_3O_5P$ | 273.155 | -dC | $C_9H_{12}N_3O_6P$ | 289.184 |
| -ddT | $C_{10}H_{13}N_2O_6P$ | 288.196 | -dT | $C_{10}H_{13}N_2O_7P$ | 304.195 |
| -ddA | $C_{10}H_{12}N_5O_4P$ | 297.210 | -dA | $C_{10}H_{12}N_5O_5P$ | 313.209 |
| -ddG | $C_{10}H_{12}N_5O_5P$ | 313.209 | -dG | $C_{10}H_{12}N_5O_6P$ | 329.208 |

In more detail, the DNA 20 is denatured (Step 1) and the two strands of DNA 20, e.g., separated by heating the native DNA to about 90°-99°C for 10 to 60 seconds. The primer 30 is then added at a concentration ranging from about, e.g., 0.1 μ M to about 10 μ M, preferably 1 μ M. The primer 30 preferably includes 15-50 nucleotide bases and is oriented such that the base 26, complementary to the base 24 adjacent the putative point mutation 10 on the target strand of DNA 22, is located at the 3' end of the primer 30. The mixture of primer 30 and target DNA 22 is cooled to between 37° to 72°C to permit the annealing and hybridizing of the primer 30 to the target DNA 22 (Step 2). The next step (Step 3) consists of adding dideoxynucleotide bases 36, 36', a polymerase, and an appropriate buffer to the mixture in order to extend the 3' end of the primer 30 bound to the target DNA 22. This step is accomplished using standard DNA polymerase technology. The base 36 which is added to the primer 30 will be complementary to the putative point mutation 10 (Step 4).

In certain embodiments, the addition of a dideoxynucleotide base 36 to the primer 30 is performed using a thermostable DNA polymerase, for example *Taq* polymerase, *AmpliTaq* FS or thermosequenase, or alternatively by using an ordinary thermolabile DNA polymerase, such as Klenow, T7, or Sequenase DNA polymerase. Preferably, thermosequenase is used. By using a thermostable polymerase, the assay, in one embodiment, involves primer extension amplification by repeating cycles of: 1) DNA denaturation by heating above the melting point of the DNA sample; 2) annealing or hybridizing the primer to the DNA by cooling to the adequate temperature; and 3) extending the primer by adding a nucleotide base. Such amplification increases the molecular ratio of the extended primer 30' over the DNA sample. This will also result in an increase in sensitivity of the assay which permits detection of a polymorphic site on a DNA sample having a picomolar concentration (pM). However, when an ordinary thermolabile DNA

- 9 -

polymerase is used, generally only one thermal cycle is possible because the activity of these enzymes is lost during the denaturation step unless fresh enzyme is added at each thermal cycle.

The number of thermal cycles necessary for primer extension amplification in any one assay is dependent on the initial concentration of the DNA sample and the sensitivity threshold of the mass spectrometer. In any one thermal cycle, the maximum increase in molar concentration of the extended primer is equal to the molar concentration of the DNA sample, assuming that all other components, i.e. primer 30 and nucleotide bases 36,36', are present in excess. The concentration of the extended primer will increase by one fold at every thermal cycle (the fold unit being the concentration of the DNA sample). In certain embodiments, where the concentration of the DNA sample is below the sensitivity threshold of the mass spectrometer, a large excess of primer over the DNA sample is used in combination with several thermal cycles in order to bring the concentration level of the extended primer above the threshold sensitivity of the mass spectrometer. At such time the molecular ratio of the primer 30 to DNA 20 ranges from about 1:1 to about 50:1 when up to 50 thermal cycles are used, and preferably from about 1:1 to about 25:1 when up to 25 thermal cycles are used.

The buffer used in combination with the polymerase is a buffer compatible with primer extension technology. However the use of a buffer containing an alkali metal generates DNA adducts which complicate the interpretation of the mass spectra, and thus renders the identification of the polymorphism more difficult. In a preferred embodiment, the buffer is a volatile acid-base combination which does not form DNA adducts. In preferred embodiments, the buffer is ammonium acetate or ammonium formate. In another preferred embodiment, the buffer is ammonium acetate in concentration ranging from 25 mM to 100 mM. In other preferred embodiments, the buffer includes a polymerase co-factor such as magnesium chloride or magnesium acetate. In a most preferred embodiment, the buffer is composed of 25-100 mM ammonium acetate with 1.5 mM magnesium acetate.

Once the primer has been extended, the identity of the added nucleotide base 36 is then determined (Step 5) by mass spectrometry by comparing the masses of the extended primer 30' and the primer 30. The mass difference between the two primers will be characteristic of the specific nucleotide base 36 which has been added to the primer 30' (see Table 1). Having determined the identity of the nucleotide base 36 on the extended primer 30', the identity of the nucleotide base at the putative point mutation 10 on the DNA 22 is then easily deduced according to the rules of Watson-Crick base pairing. In one embodiment, mass spectrometry analysis is carried out using electrospray, matrix assisted laser desorption ionization (MALDI), fast atom bombardment ionization (FAB) or plasma desorption ionization. In other embodiments, the mass spectrometry analysis is carried out using time-of-flight (TOF), quadrupole, ion

- 10 -

trap and sector analysis modes. In a preferred embodiment, mass spectrometry analysis is carried out using MALDI-TOF mass spectrometry.

Referring to Fig. 2, the results of a single nucleotide polymorphism (SNP) assay using primer extension reaction with polymerase is shown. For this experiment, the target DNA consisted of the sequence *lacI* (wild type "C") 5'-CTGAATTACA TTCCCAACCG CGTGGCACAA CAACTGGCGG GCAAACAGTC GTTGCTGATT-3' (SEQ ID NO:1), where C represents the putative point mutation 10. For the purpose of primer extension amplification, thermosequenase, a genetically altered form of T7 DNA polymerase, from Amersham Life Science (Arlington Heights, Illinois, USA) was used to repeatedly extend the primer 3'- TTGACCGCCC GTTTGTTCAGC-5' (SEQ ID NO:2). The primer was present in 2.5X molar excess over the target DNA and it was determined that about one half of the primer was converted in a single thermal cycle of 95°C for 10 sec., 37°C for 60 sec., and 72°C for 60 sec. Substantially all of the primer was extended in three cycles. As is shown in Fig 2, when dideoxyadenosine triphosphate (ddATP) (line A), and dideoxycytidine triphosphate (ddCTP) (line B) were individually added as the base 36', substantially none of the added dideoxynucleotide base was incorporated into the primer 30. When dideoxyguanosine triphosphate (dGTP) (line C) was added as the base, the primer 30 was all converted to the extended primer 30' having an additional dideoxyguanylate (ddG) base. When dideoxythymidine triphosphate (dTTP) was added as the base (line D) a small portion of the dideoxythymidylate base (ddT) was non specifically added to the primer. However, when all four dideoxynucleotides (line E) were added together, again only the dideoxyguanylate was incorporated into the primer.

Referring to Fig. 3, when the method is used to investigate single nucleotide polymorphisms in a heterozygous sample, the DNA strands 22, 22' of the heterozygous sample contain each a different nucleotide base 10 and 10', respectively, at the putative point mutation. In the example shown, the sample contain a 1:1 mixture of DNA target having cytosine or thymine at the putative point mutations 10, 10' respectively. As discussed above the DNA is denatured and each of the target strands 22, 22' hybridized with a primer 30 (Step 1). The four dideoxynucleotide bases are then added along with the polymerase and buffer, as discussed above. The dideoxynucleotide bases 36, 36'' which are complementary to the putative point mutations 10, 10', respectively, thereby extend the primer 30 (Step 2). In the example shown, the primer 30 hybridized with the target DNA 22 and 22' is extended by dideoxyguanylate (ddG) 36 and dideoxyadenylate (ddA) 36'', respectively. Referring to Fig. 4, when analyzed by mass spectroscopy as described above, the result is a mass spectrum having two peaks which differ by a mass of 16.050 Da which is close to the expected molecular weight difference of 15.999 Da between the two dideoxynucleotide bases dG and dA which extend each primer 30' and 30'' and which is within the experimental error of the mass spectrometer.

- 11 -

In another embodiment and referring to Fig. 5, the primer 56 includes a mass tag 54. The mass tag 54 is preferably any form of mass extension to the primer 56 which does not hybridize with the target DNA 60 and which also does not interfere with the stability of the hybrid or interfere with the extension of the primer using DNA polymerase. Thus in one embodiment the mass tag 54 is constructed of two or more nucleotide bases 54, added to the 5' end of the primer 56, which are not complementary to the target DNA 60. For convenience of synthesis, the mass tag 54 may consist of a homopolymer tail 54 synthesized to the 5' end of the primer 56. In one embodiment the homopolymer tail 54 is a poly-deoxythymidylate, T_n wherein n is an integer ranging from 1 to 40.

The purpose of the mass tag is to separate the primers and the extended primers which might otherwise be unresolvable by mass spectrometry. Thus, two primers or extended primers may have the same number of bases and hence have approximately the same molecular weight and therefore be unresolvable by mass spectrometry, the addition of a mass tag to one of the primers adds sufficient mass to permit them to be resolved. This permits two or more primers to be used in the same assay to probe for multiple single nucleotide polymorphisms simultaneously.

To perform such an assay, probing for two single nucleotide polymorphisms for example, two primers 30, 56 are provided (Step 1) which are each respectively complementary to the bases adjacent the two different single nucleotide polymorphisms 10, 50, shown on a single target strand of DNA 60. The primers 30, 56 differ in mass by at least the addition of a mass tag 54 to one of the primers 56. As described previously, each primer 30, 56 is hybridized to its complementary region of the target DNA 60 and the dideoxynucleotide bases 36, polymerase and buffer (not shown) are added to the mixture (Step 2). Each dideoxynucleotide base 64, 68 which is complementary to its respective single nucleotide polymorphism 10, 50 then elongates the primer 30, 56 respectively (Step 3) as described above. The target DNA 60 and the primers 30, 56, which have been elongated by the bases 64, 68 respectively, are analyzed by mass spectrometry as described previously.

The multiple SNP assay applies equally well if the SNPs are located on the same DNA segment but different loci, or on two or several DNA segments. The several DNA segments may be complementary strands or have fully distinct nucleotide sequences. Some assays may be designed such that the primers hybridize on either one or the other of the two complementary DNA strands to be probed. This is particularly well suited to multiplex assay experiments where two SNPs are separated from each other by a number of nucleotide bases smaller than the number of nucleotide bases of the downstream primers thereby avoiding overlapping of their binding site on the DNA. Since the 3' ends of the sequences of two complementary DNA strands run opposite to each other, it is possible to design two primers to probe two

- 12 -

SNPs close to each other so that one primer binds on one strand of DNA and the downstream one binds on the other strand, each having their 5'-end extending away from the two putative point mutations and their 3' end adjacent one putative point mutation and such that the primers have no complementary segments in their sequences.

5 The results of such multiple single nucleotide polymorphism assays are shown in Figs. 6a, 6b, 6c, 6d, 6e and 6f. In Fig. 6a and 6b, the experiment depicts five primers (R377, R400, R421, R447, R355) designed to examine five loci on the breast cancer susceptibility gene *BRCA1*. As shown in Table 2 below, the five primers are each complementary to a different region of the gene. Each primer has the same core length of eighteen complementary bases but each primer but one has a different mass tag from the others; one primer having no mass tag. Therefore, each primer has a significantly different mass from all the other primers. The mass spectra of the five primers before extension are shown in Fig. 6a. When the primers are extended using the method described above with twenty five thermal cycles and the mass spectra taken, the mass of each primer is increased by the mass of the added nucleotide. Note that primer R355 was entirely converted by the addition of dideoxyguanine and as such there is no peak in Fig. 6b corresponding to the unelongated primer. Unexpectedly, the addition of a decadeoxythymidylate residue (dT₁₀) on the 5' end of R355 did not appear to interfere with the extension of primer R377 even though the two primers anneal to the DNA within a couple of nucleotide bases.

 Figs. 6c and 6d depict an experiment with seven primers designed to examine seven loci on the breast cancer susceptibility gene *BRCA1* (R355, R377, R400, R421, R447, R480 on one strand and F320 on the other strand). All the primers, shown in Fig. 6c, were extended by the expected nucleotide bases within twenty five thermal cycles. In Fig. 6d, all fourteen peaks of the seven primers and the seven extended primers were resolved, but two primers, R355 and F320, were nearly quantitatively extended. Note that primer F320 gave two products: one corresponding to the addition of ddT and the other corresponding to the addition of dT and ddC. T and C matches the two nucleotides bases downstream of the binding site of primer F320 on *BRCA1*. This two-base extension is due to carry-over of residual dcoxynucleotide bases from PCR amplification of target DNA *BRCA1*. Residual dcoxynucleotide bases can be eliminated by purification of DNA target after PCR

Table 2: SNP Primers for Multiplex Assays

| Targets | Core Length (# of bases) | Mass Tag (# of bases) | Total Length (# of bases) | Mass (amu) | Base Extended |
|---------|-----------------------------|--------------------------|------------------------------|---------------|------------------|
| BRCA1 | | | | | |
| R377 | 18 | 0 | 18 | 5,477.58 | ddG |
| F320 | 20 | 0 | 20 | 5,996.94 | ddT |
| R400 | 18 | 4 | 22 | 6,594.34 | ddT |
| R421 | 18 | 8 | 24 | 7,393.84 | ddC |
| R480 | 25 | 0 | 25 | 7,696.04 | ddA |
| R447 | 18 | 8 | 26 | 8,112.34 | ddA |
| R355 | 18 | 10 | 28 | 8,599.61 | ddG |
| lacz | | | | | |
| b2-1 | 15 | 0 | 15 | 4,534.97 | ddG |
| b2-1 | 20 | 0 | 20 | 6,059.95 | ddG |

| Targets | SEQ ID NOs | Nucleotide Sequence |
|---------|-------------|--------------------------------------|
| BRCA1 | | |
| R377 | SEQ ID NO:3 | 5'-TGCTTTGTTC TGGATTTC-3' |
| F320 | SEQ ID NO:4 | 5'-AGCCTTCTAA CAGCTACCCT-3' |
| R400 | SEQ ID NO:5 | 5'-TTTTACAATA CACACCTTTT TC-3' |
| R421 | SEQ ID NO:6 | 5'-TTTTTTTAAG ATATCAGTGT TTGG-3' |
| R480 | SEQ ID NO:7 | 5'-CATAAAATGT TGGAGCTAGG TCCTT-3' |
| R447 | SEQ ID NO:8 | 5'-TTTTTTTAT AAAGGGAAG GAAAGA-3' |
| R355 | SEQ ID NO:9 | 5'-TTTTTTTTTT GTCCTCAAGG GCAGAAGA-3' |

Table 2:(continued)

| Targets | SEQ ID NOs | Nucleotide Sequence |
|-------------|--------------|-----------------------------|
| <i>lacZ</i> | | |
| b2-1 | SEQ ID NO:10 | 5'-GTTTGCCCGC CAGTT-3' |
| b2-1 | SEQ ID NO:2 | 5'-CGACTGTTTG CCCGCCAGTT-3' |

amplification by ethanol precipitation, gel filtration, or incubation with shrimp alkaline phosphatase.

- 14 -

In Fig. 6e and 6f, the experiment depicts seven primers designed to examine five loci on the breast cancer susceptibility gene *BRCA1* (R355, R377, R400, R421, R447) and one locus on a 60 base synthetic sequence of the *lacZ* gene (b2-1 with a 15-mer and a 20-mer). Two primers hybridizing to the same site (b2-1) but differing in length by five bases were included to determine the effect of the primers' core length on their ability to hybridize with their target site and therefore their ability to undergo nucleotide base extension. All seven primers, shown in Fig 6e, were extended by the expected nucleotide base. Three of the primers were quantitatively extended within twenty five thermal cycles. In Fig. 6f, all fourteen peaks of the seven primers and the seven extended primers were resolved, but three primers (both b2-1 and R355), were quantitatively extended, while two primers, R400 and R421, produced significantly lower amounts of extended primers than the other primers. Note that both primers b2-1 gave also two extended primers each: the expected nucleotide base ddG and a two-base extension, dGddT, matching to the sequence of the *lacZ* gene downstream of the binding site of the primers b2-1.

In another embodiment, the present invention relates to a method for the identification of single nucleotide polymorphism using ligase chain reaction assay (LCR) on either a single stranded or a double stranded DNA. Referring to Fig. 7, a set of ten oligonucleotide probes 70, 71, 71', 71'', 71''', 72, 73, 73', 73'' and 73''' are provided to the two strands of a DNA 20, 22 having a single putative point mutation 9, 10. In the first cycle, the oligonucleotide probes are hybridized at the putative point mutation 9, 10 on each strand of DNA 20, 22 such that probes 70 and 71 together overlap the putative mutation site 9 on DNA strand 20 and probes 72 and 73 together overlap the putative mutation site 10 on DNA strand 22 (Step 1). Each set of probes 70, 71, and 72, 73 on each DNA strand 20, 22 is then ligated (Step 2) using a ligase and appropriate buffer. In the successive cycles, the ligated probes 80 and 81 will serve as template along with DNA strands 20 and 22 to ligate the probes 70, 71, 72 and 73 resulting in an amplification of the genetic information contained in the segment of DNA 20,22 expressed in the ligated probes 80 and 81. The resulting two ligated oligonucleotide probes 80, 81 are analyzed by mass spectrometry. The ligation step will produce ligated probes 80 and 81 only if the probes 71 and 73 which match the putative point mutation 9, 10 are hybridized to their respective DNA strand. If any of the probes which are only partially hybridized to the DNA in the region of the putative mutation, such as probes 71', 71'', 71''', 73', 73'' and 73''', then the ligase will not produce a ligated probe. Therefore, the identity of the bases 76 and 78 of the ligated probes 80 or 81 matching the putative point mutation 9, 10 is deduced by measuring their mass and comparing it with the sum of the masses of 70 and 71, 70 and 71', 70 and 71'' and 70 and 71''' for ligated probe 80 and the sum of the masses of 72 and 73, 72 and 73', 72 and 73'' and 72 and 73''' for ligated probe 81. Once the identity of the bases 76, 78 in the ligated probes 80, 81 is determined, the identity of the bases at the putative point mutation 9, 10 is deduced according to the rules of Watson-Crick base pairing.

- 15 -

Note that the information provided by ligated probe 81 is redundant with the information provided by the ligated probe 80 which allows one to deduce the identity of the base at the putative mutation site on the strand of DNA 20. The identity of the base on the second strand 22 can be deduced from the identity of the base on the strand 22 also according to the rules of Watson-Crick base pairing. This redundancy provides for greater reliability in the determination of the assay by providing a second set of data to analyze.

When using a single stranded DNA 20 with LCR, the first cycle will produce only one ligated probe 80 from probes 70 and 71 having a complementary sequence. Then, in the successive cycles, ligated probe 80 and single strand 20, both will serve as a template for the ligation of probes 70, 71, 72 and 73 to produce ligated probes 80 and 81. The subsequent cycles will operate in an amplification of the genetic information as if a double stranded DNA was used initially by producing large amounts of both ligated probes 80 and 81.

In one embodiment, a mass tag is added to one of the probes, for example, to form a tagged probe (not shown). The mass tag may be added to either the 3' end or the 5' end of a probe so long as it is added on the end that does not participate in the ligation between two probes. The number of bases in the resulting ligated probe (not shown) is now increased by the number of bases added in the tag of primer. This mass tagging increases the difference in masses of ligated probes and thus allow a greater resolution in the results of a ligation assay where the masses of ligated probes otherwise would be too close to permit a good resolution of their respective peak by mass spectrometry. This mass tagging may also be used to facilitate the interpretation of the assay experiment by differentiating between primers 71, 71', 71'' and 71''' or between primers 73, 73', 73'' and 73'''.

Referring to Fig. 8, the results of a single nucleotide (SNP) assay using ligase reaction on a homozygous DNA sample analyzed by mass spectrometry is shown. For this experiment, the target DNA had the following nucleotide sequence:

5'-CTGAATTACA TTCCCAACCG CGTGGCACAA CAACTGGCGG GCAAACAGTC
GTTGCTGATT-3' (SEQ ID NO:1)

wherein C is the point mutation. Two constant probes and eight variable probes were prepared. The sequences of the probes (shown in Table 3) were designed such that a nineteen-base probe with a twenty one-base probe were ligated to form a forty-base ligated oligonucleotide probe and a twenty-base probe with a twenty one-base probe were ligated to form a forty one-base ligated probe.

Table 3: Probes for Ligase assay

| Probe | Length | Sequence | SEQ ID NOs | Mass (Da) |
|----------|--------|--|--------------|---|
| constant | 19 | 5'-TTGTGCCACGCGGTTGGGA-3' | SEQ ID NO:11 | 5955.808 |
| variable | 21 | 5'-CGACTGTTTGCCCGCCAGTTN-3' ⁽¹⁾ | SEQ ID NO:12 | 6,059.931 + m _x ⁽²⁾ |
| variable | 21 | 5'-TTCCCAACCGCGTGGCACAAN-3' ⁽¹⁾ | SEQ ID NO:14 | 6,046.949 + m _x ⁽²⁾ |
| constant | 20 | 5'-AACTGGCGGGCAAACAGTCG-3' | | 6256.035 |

(1) N represents any of A, T, C or G, (2) m_x represents the mass of nucleotide N.

The expected mass for each ligated probes (with G/C at the variable base) is 12,326.93 Da and 12,574.15 Da. As can be seen from Fig. 8, the forty-base and forty one-base ligated probes are easily distinguished from one another with ~248 Da in mass difference. Referring to Fig. 9a, the results of a SNP assay using ligase on a heterozygous DNA sample, having both C/T and T/A at the point mutation, analyzed by mass spectrometry is shown. This produced two forty one-base ligated probes with 15.999 Da difference in mass due to G and A, and two forty-base ligated probes with 15.011 Da difference in mass due to C and T. Referring to Fig.9b, a highly schematic diagram of an SNP assay using ligase on an heterozygous DNA sample is shown.

In another embodiment, the present invention relates to a method for the identification of single nucleotide polymorphism using oligonucleotide ligation assay (OLA) on only one DNA strand 20 having a putative point mutation 9. The method consists of providing a set of five probes 70, 71, 71', 71'' and 71''', one constant 70 and four variable probes 71-71''', wherein the constant probe and only one of the four variable probes hybridize on the DNA 20 such that the two hybridizing probes overlap the putative mutation site and form a perfect match with the base at the putative point mutation 9 of the DNA 20; ligating the two hybridized probes using oligonucleotide ligation assay; identifying the ligated probes by mass spectrometry in a manner similar to the assay using ligation chain reaction with the double stranded DNA.

In yet another embodiment, the present invention includes combinatorial methods for the identification of the nucleotide bases at a putative mutation site that results from either a deletion, an insertion or a substitution of several nucleotide bases by probing more than one nucleotide base at the putative mutation site. Generally, the combinatorial method includes the steps of producing a combination of extended primers of varying length, the combination of extended primers being a ladder sequence defining mixture of short DNA segments where each step of the ladder sequence represents a base at the putative mutation site, and analyzing the combination of extended primers by mass spectrometry (where the

- 17 -

mass of each extended primer of specific length defining a step of the ladder sequence). The identity of the sequence at the putative mutation site is deduced by the sequential determination of the mass differences between each step of the ladder sequence starting either from the longest extended primer and proceeding to the initial primer used in the assay or starting from the initial primer and proceeding to the longest extended primer.

In more detail and referring to Fig. 10, one embodiment of the combinatorial methods consists of first hybridizing a primer 30 having a sequence complementary to a target DNA segment 22 having a putative mutation site 11 (Step 1). The 3'-end of the primer 30 is positioned adjacent the putative mutation site 11. Next, primer 30 is extended by adding a multiplicity of nucleotide bases 36, 38 on primer 30 (Step 2). For this, each of the possible complementary bases in form of deoxynucleotides 38 and dideoxynucleotides 36, is added to the mixture along with a DNA polymerase and appropriate buffer solution (not shown). Extended primer 30', 42, 42' or 42'' is then melted from DNA 22 (Step 3), and DNA 22 is then recycled with excess primer 30 to step 1. At each thermal cycle (Steps 1, 2 and 3), the polymerase extends the primer 30 with the available nucleoside bases 38 and 36 by matching the added bases to the bases of the putative mutation site 11, the DNA 22 serving as a template, until the polymerase either incorporates in the growing primer a stopping base (dideoxynucleotide 36); or the polymerase runs out of available nucleotide bases 38. Thus, by providing a mixture of all the nucleotide bases 36 and 38, the process will generate at each thermal cycle an extended primer 42 of random length. By repeating the thermal cycle a number of times, the process generates a combination of extended primers 30', 42, 42' and 42'' of varying length where the sequence of the extended portion 15 of the extended primers is made of deoxynucleotide bases plus the terminal dideoxynucleotide base, all of which match the sequence of the putative mutation site 11 of DNA 22.

Preferably, a thermostable DNA polymerase is used, however, any DNA polymerase can be used provided that, if amplification is required, additional amount of polymerase is added before each thermal cycle. A thermal cycle includes a hybridizing step conducted preferably at 37 °C for 60 seconds; a primer extension step; and a denaturing step where the DNA sample is heated to a temperature preferably 95 °C for 10 seconds sufficient to melt any DNA/DNA (20,22) helix, any DNA/primer hybrid (22, 30) or any DNA/extended primer (22, 42) duplexes. The ratio of dideoxynucleotides bases 36 to deoxynucleotide bases 38 (ddNTP:dNTP) for each base may be varied from 1,000:1 to 0.025:1. The preferred range is from about 2:1 to 0.05:1 when thermosequencase is used.

The combination of extended primers 30', 42, 42' and 42'' creates a ladder sequence defining mixture wherein each extended primer 30', 42, 42' and 42'' represents a step of the ladder sequence. Each

- 18 -

extended primer 30', 42, 42' and 42'' is terminated by one of the four dideoxynucleotide bases 36. The mass difference between each step of the ladder sequence therefore corresponds to the mass of one of the four dideoxynucleotide bases 36. The identity of the sequence at the polymorphic site is then determined by measuring the difference in mass between two successive steps of the ladder sequence, correlating the mass to one of the four dideoxynucleotide bases and repeating this for each step of the ladder sequence (Step 4). Once the added sequence 15 on the extended primers is identified, the identity of the sequence at the putative mutation site 11 is deduced by following the rules of Watson-Crick base pairing.

Referring to Fig. 11, the results of a combinatorial assay conducted on an oligonucleotide target sequence: 5'-CTGAATTACA TTCCCAACCG CGTGGCACA CAACTGGCGG GCAAACAGTC GTTGCTGATT-3' (SEQ ID NO:1) where GTGGCACAAC (portion of SEQ ID NO:1) represents the putative mutation site 11. The twenty-base long sequence: 3'-TTGACCGCCC GTTTGTCAGC-5' (SEQ ID NO:2) is used as the primer.

The thermal cycles were conducted on 0.4 μ M of oligonucleotide target sequence in presence of 2 μ M of primer 30, 20 μ M of each triphosphate base (dideoxy and deoxy combined), with a ratio ddNTP:dNTP of 0.1, and 30 units/mL of thermosequenase in 25 mM ammonium acetate at pH 9.3 and 2mM magnesium chloride buffer solution. The reaction mixture was thermally cycled at 95 °C for 10 sec., 37 °C for 30 sec. and 72 °C for 60 sec. for 20 cycles in a Perkin Elmer 9600 DNA Thermal Cycler. An amount of 10 μ L of the reaction mixture was then sampled out and applied through a pipette tip packed with a few microliters of POROS[®] R1, washed with a 100mM triethylammonium acetate solution and eluted in about 2 μ L of acetonitrile: water (80:20) solution. The resulting solution was then analyzed by MALDI-TOF mass spectrometry. As seen in Fig. 11, all of the initial primer 30 was extended within the twenty cycles and no peak corresponding to the initial primer 30 is observed in the mass spectrum. The mass spectrum displays the mass of ten extended primers from primer with twenty one bases to primer with thirty bases. The sequence of the added portion 15 is deduced by measuring the mass difference between each step of the ladder sequence and is shown to be 5'-GTTGTGCCAC-3' (SEQ ID NO:15). Thus, the sequence of the putative mutation site 11 is deduced as being the complementary sequence: 3'-CAACACGGTG-5' (portion of SEQ ID NO:1).

In another embodiment, and referring to Fig. 12, the combinatorial method includes hybridizing a primer 30 having a sequence complementary to a target DNA 22 segment having a putative mutation site 11 (Step 1). The 3' end of the primer 30 is adjacent putative mutation site 11. Primer 30 is extended (Step 2) by adding a full set of deoxynucleotide bases 38 with only one of the four dideoxynucleotide bases 36, ddG for example, along with a DNA polymerase and appropriate buffer solution (note shown). The

- 19 -

mixture is subjected to the thermal cycle to amplify primer 30 into measurable amounts of extended primers 43, 43' and 43''. Again, the last base of each extended primer is the dideoxynucleotide base 36. Since the mass difference between each extended primers 43, 43' and 43'' is most likely going to be greater than the mass of one dideoxynucleotide base 36, no sequencing data may be extracted at this stage of the process. The mixture of extended primers 43, 43' and 43'' is then subjected to digestion by an exonuclease (Step 4) to create a ladder sequence defining mixture of extended primers of varying length comprising extended primers 43, 43' and 43'' and newly formed digested primers 44, 44', 44'', 45, 45', 45'', 46 and 46'. The addition of a base 36 is to control the length of the extended primers, otherwise the primer 30 would be extended for a long and undefined length and its molar concentration would be too low to be detected by mass spectrometry. By limiting the primer extension to a relative short length, a high molar concentration of the extended primers may be attained. Also, by producing a controlled distribution in length of the extended primers before the digestion step, a controlled distribution in length of the digested primers may be attained with detectable molar concentrations. The combination of extended primers and digested primers defines a ladder sequence where the mass difference between extended primer 43 and first digested primer 44, for example, is the mass of dideoxynucleotide base 36 and the mass difference between each successive step of the ladder sequence formed by digested primers 44 and 44' or 44' and 44'', for example, is the mass of one of the four deoxynucleotide base 38. To facilitate the reading of the mass spectrum resulting from the digestion, it is preferable to record a mass spectrum of the mixture of extended primers 43, 43' and 43'' before effecting the digestion to help locate their peaks in the mass spectrum resulting from the digestion. Once every step of the ladder sequence has been attributed to either a base 36 or 38 (Step 5), the sequence of the putative mutation site is deduced by following the rules of Watson-Crick base pairing. This method is particularly well suited when the number of bases to be determined at the putative mutation site 11 is greater than ten.

The digestion of the mixture of extended primers 43, 43' and 43'' may be carried with any known exonuclease such as phosphodiesterase type I, exonuclease I, exonuclease III, exonuclease V, exonuclease VII, and DNA polymerase III.

In still yet another embodiment, the present invention provides a multiplex combinatorial method for the identification of multiple putative mutation sites that result from a deletion, an insertion or a multiple substitution of a nucleotide base. In referring to Fig. 13, the methods include hybridizing a multiplicity m of primers 30 and 58, each having a sequence complementary and to a putative mutation site 11 and 11' respectively, of a DNA 62 having more than one putative mutation site (Step 1). Next the method extends the multiplicity m of primers in the presence of a multiplicity of nucleotide bases 36 and 38 with a DNA polymerase and appropriate buffer solution (not shown) such that each primer is extended by

- 20 -

more than one nucleotide base (Step 2). The multiplicity of extended primers 30' and 58 is melted from DNA 62 (Step 3) and DNA 62 is recycled (Step 1) to hybridize with excess of primers 30 and 57. At each cycle, the polymerase extends the multiplicity of primers 30 and 57 with the available nucleotide bases 36 and 38 by matching the added bases at the polymorphic sites 11 and 11', the DNA 62 serving as a
5 template, producing a multiplicity of extended primers 53' 42, 42'', 58, 58', 58'' and 58'''. Finally, the combination of extended primers is identified by mass spectrometry (Step 4). The combination of extended primers defines a series of ladder sequences, each ladder sequence being an imprint of each putative mutation site 11 and 11'. The sequence of each putative mutation site is then determined by subtracting the masses between two adjacent steps of each ladder sequence and attributing the mass difference to a
10 nucleotide base. The base at the putative mutation site is then deduced according to the rules of Watson-Crick base pairing. This is repeated for every step of a ladder sequence.

In still another embodiment, the multiplex combinatorial method includes the combination of primer extension and digestion. First, a multiplicity m of primers, each having a sequence complementary and adjacent a putative mutation site of a DNA from a sample having more than one putative mutation site
15 is hybridized to the DNA. Next, the multiplicity m of primers is amplified in the presence of a multiplicity of deoxynucleotide bases and at least one dideoxynucleotide base with a polymerase and appropriate buffer solution so that each primer is extended by more than one nucleotide base. Third, the multiplicity of extended primers is subjected to a digestion using an exonuclease. Finally, the combination of extended and digested primers is analyzed by mass spectrometry. The sequence of each putative mutation site is
20 then determined as described earlier. To attain a good resolution of the mass spectrum obtained from the combination of extended primers, it is preferred to use a multiplicity m of primers wherein $(m-1)$ primers have a mass tag on their 5' end.

In yet other embodiments, the methods of present the invention relates to the identification of nucleotide polymorphisms in RNA target sequences. Any of the methods described above can be applied to
25 determine polymorphism in an RNA by substituting an RNA target in place of a DNA target and reverse transcriptase in place of DNA polymerase. In a preferred embodiment, thermal cycles are performed using a thermostable reverse transcriptase, such as thermus thermophilus (rTth) DNA polymerase for example.

A highly schematic diagram of an instrument used to perform the various methods of the invention is depicted in Fig. 14. Each of the primers to be used are added separately to a sample plate 90 which
30 includes a series of sample locations; one for each primer. The sample plate 90 is placed in a MALDI-TOF spectrometer 100 which determines the mass of each of the individual primers used in the sequencing. Next the hybridized modified primer and target polynucleotide is added to a sample plate 90' and this

- 21 -

sample plate 90' is placed in the MALDI-TOF spectrometer 100. The mass of the modified primer and the mass of the target polynucleotide are then determined; the hybridization being destroyed by the laser light during the desorption ionization process which occurs within the MALDI-TOF 100.

5 A processor 120 which may be a general purpose computer or a specialized processor collects the mass data from the MALDI-TOF 100 and places the data into memory 124 and optionally onto a disk or other off-line storage 126. The processor 120 then determines the mass difference between each primer and its corresponding modified primer. The processor 120 then compares the mass difference for each primer with a table of nucleotide masses stored on disk 126 and read into memory 124. From this tabular comparison of the mass difference of the primer and the modified primer with the nucleotide base masses,
10 the nucleotide additions to the primer is then determined. The results are either displayed to the user 130 or stored on disk 126 for later analysis.

Briefly, and referring to Fig. 15, an embodiment of an algorithm to determine sequence data includes the step of taking the mass data from the MALDI-TOF 100 (Step 150) and from the data determining the mass of each of the primer used (Step 154) and each modified primer formed (Step 158).
15 The table of nucleotide masses located in memory 124 is compared (Step 162) to the mass differences determined in the previous step. By knowing the individual masses of the nucleotides, the nucleotide sequence is generated (Step 166). Once the data and sequence have been determined, the data and/or the sequence are displayed (Step 170) on the user's I/O device such as a terminal 130 or are stored in off line storage 126.

20 Having described preferred embodiments of the invention, it will now become apparent to one of skilled in the art that other embodiments incorporating the concepts disclosed herein may be used. It is felt, therefore, that these embodiments should not be limited to disclosed embodiments but rather should be limited only by the spirit and scope of the following claims.

- 22 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: HAFF, Lawrence A.
SMIRNOV, Igor P.
- (ii) TITLE OF INVENTION: METHODS FOR DETERMINING SEQUENCE
INFORMATION IN POLYNUCLEOTIDES USING MASS SPECTROMETRY
- (iii) NUMBER OF SEQUENCES: 15
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Patent Administrator, Testa Hurwitz &
Thibeault, LLP
 - (B) STREET: 125 High Street
 - (C) CITY: Boston
 - (D) STATE: Massachusetts
 - (E) COUNTRY: USA
 - (F) ZIP: 02110
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/726,090
 - (B) FILING DATE: 04-OCT-1996
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: TURANO, THOMAS A.
 - (B) REGISTRATION NUMBER: 35,722
 - (C) REFERENCE/DOCKET NUMBER: SYP-123
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (617) 248-7000
 - (B) TELEFAX: (617) 248-7100

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTGAATTACA TTCCAACCG CGTGGCACAA CAACTGGCGG GCAAACAGTC
GTTGCTGATT 60

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

- 23 -

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGACTGTTTG CCCGCCAGTT

20

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TGCTTTGTTC TGGATTTC

18

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGCCTTCTAA CAGCTACCCT

20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTTTACAATA CACACCTTTT TC

22

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 24 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
TTTTTTTAAG ATATCAGTGT TTGG 24

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
CATAAAATGT TGGAGCTAGG TCCTT 25

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
TTTTTTTTTAT AAAGGGAAG GAAAGA 26

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
TTTTTTTTTTT GTCCTCAAGG GCAGAAGA 28

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

- 25 -

GTTTGCCCGC CAGTT

15

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTGTGCCACG CGGTTGGGA

19

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGACTGTTTG CCCGCCAGTT N

21

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTCCCAACCG CGTGGCACAA N

21

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AACTGGCGGG CAAACAGTCG

20

(2) INFORMATION FOR SEQ ID NO:15:

- 26 -

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTTGTGCCAC

10

CLAIMS

What is claimed is:

- 1 1. A method for identifying at least one putative point mutation in at least one polynucleotide using mass
2 spectrometry, comprising the steps of:
 - 3 a. hybridizing at least one primer to a single strand of said at least one polynucleotide having
4 at least one putative point mutation, the 3'- end of said at least one primer being hybridized
5 adjacent the at least one putative point mutation;
 - 6 b. extending the 3'- end of said at least one primer by adding a single nucleotide base
7 complementary to the at least one putative point mutation of the at least one polynucleotide
8 using a polynucleotide polymerase having a relative rate of incorporation for
9 dideoxynucleotides of the same magnitude than the rate of incorporation of deoxynucleotides,
10 and at least one dideoxynucleoside triphosphate base (ddNTP), thereby forming at least one
11 extended primer; and
 - 12 c. analyzing the added nucleotide base on the at least one extended primer using mass
13 spectrometry to determine the identity of the nucleotide base at the at least one putative point
14 mutation.
- 1 2. The method of claim 1, wherein step a. comprises
2 hybridizing a plurality of primers to at least one single strand of said at least one
3 polynucleotide having said plurality of putative point mutations, the 3'- end of each of said
4 plurality of primers being hybridized adjacent a respective one of said plurality of putative
5 point mutations;
6 step b. comprises
7 extending the 3'-end of said plurality of primers by adding at least one of a plurality of
8 nucleotide bases whereby at least one of said plurality of nucleotide bases is added to each
9 respective one of said plurality of hybridized primers complementary to said respective one
10 putative point mutation;
11 and step c. comprises
12 analyzing the added nucleotide base on each extended primer using mass spectrometry to
13 determine the identity of the nucleotide base at each putative point mutation.
- 1 3. The method of claim 1 or 2, wherein said primer comprises a mass tag.

- 1 4. The method of claim 1 or 2, wherein said polymerase is provided in a buffer volatile under conditions
2 present in a mass spectrometer.
- 1 5. The method of claim 1 or 2, wherein steps a) and b) are repeated for a sufficient number of times to
2 detect each extended primer.
- 1 6. A method for determining sequence information in at least one region of at least one polynucleotide
2 using mass spectrometry, comprising the steps of:
- 3 a. hybridizing at least one primer to a single strand of said at least one polynucleotide having
4 at least one region of interest, the 3'- end of said at least one primer being hybridized adjacent
5 said at least one region of interest; and
- 6 b. extending the 3'- end of said at least one primer thereby forming at least one extended
7 primer, said extending step comprising providing a polynucleotide polymerase having a relative
8 rate of incorporation of dideoxynucleotides of the same magnitude than the rate of
9 incorporation of deoxynucleotides and at least one nucleotide base selected from the group
10 consisting of deoxynucleosides triphosphate (dNTP) and dideoxynucleosides triphosphate
11 (ddNTP), provided that at least one of said ddNTP bases is present, thereby forming at least
12 one extended primer having at least one added nucleotide base; and
- 13 c. analyzing the added at least one nucleotide base on the at least one extended primer using
14 mass spectrometry to determine the identity of the nucleotide bases in the at least one region of
15 interest.
- 1 7. The method of claim 6, wherein step a. comprises
2 hybridizing a plurality of primers to at least one single strand of said at least one
3 polynucleotide having a plurality of regions of interest, the 3'- end of each of said plurality of
4 primers being hybridized adjacent a respective one of said plurality of regions of interest;
5 step b. comprises
6 extending the 3'- end of said primers by adding a plurality of nucleotide bases to said plurality
7 of hybridized primers, whereby at least one of said plurality of nucleotide bases is added to
8 each respective one of said plurality of hybridized primers complementary to said respective
9 one of said regions of interest, thereby forming a plurality of extended primers each having at
10 least one added nucleotide base, comprising providing a plurality of said nucleotide bases
11 and step c. comprises

12 analyzing the added nucleotide bases on each extended primers using mass spectrometry to determine
13 the identity of the nucleotide bases in each of the regions of interest.

1 8. The method of claim 6 or 7, wherein said primer comprises a mass tag.

1 9. The method of claim 6 or 7, wherein said polymerase is provided in a buffer volatile under conditions
2 present in a mass spectrometer.

1 10. The method of claim 6 or 7, wherein steps a) and b) are repeated for a sufficient number of times to
2 produce a plurality of extended primers defining at least one ladder sequence complementary to said at
3 least one region of interest.

1 11. The method of claim 6 or 7, further comprising the step of:
2 providing an exonuclease to the mixture resulting from step b) thereby producing a plurality of
3 extended primers defining at least one ladder sequence complementary to said at least one region of
4 interest.

1/20

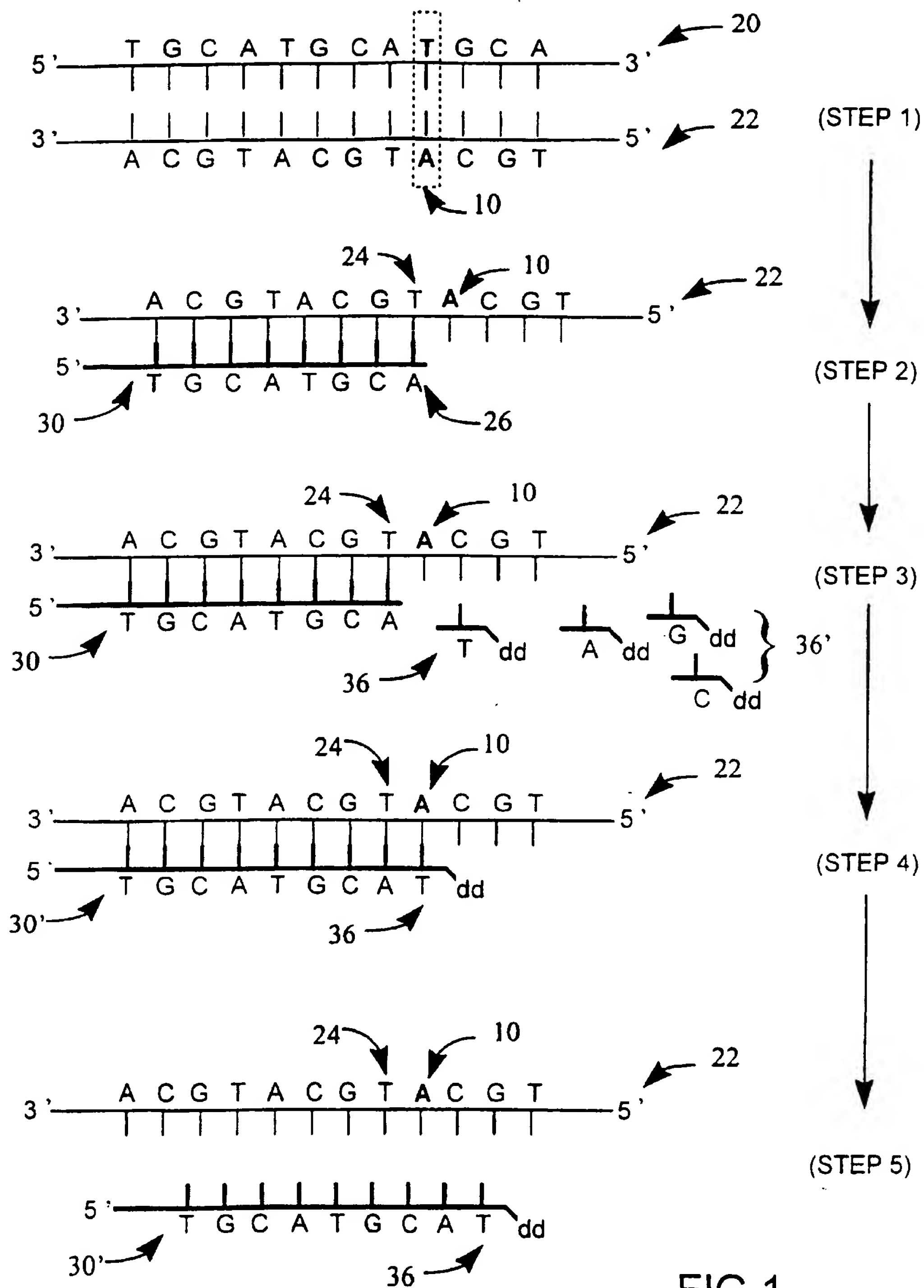


FIG.1

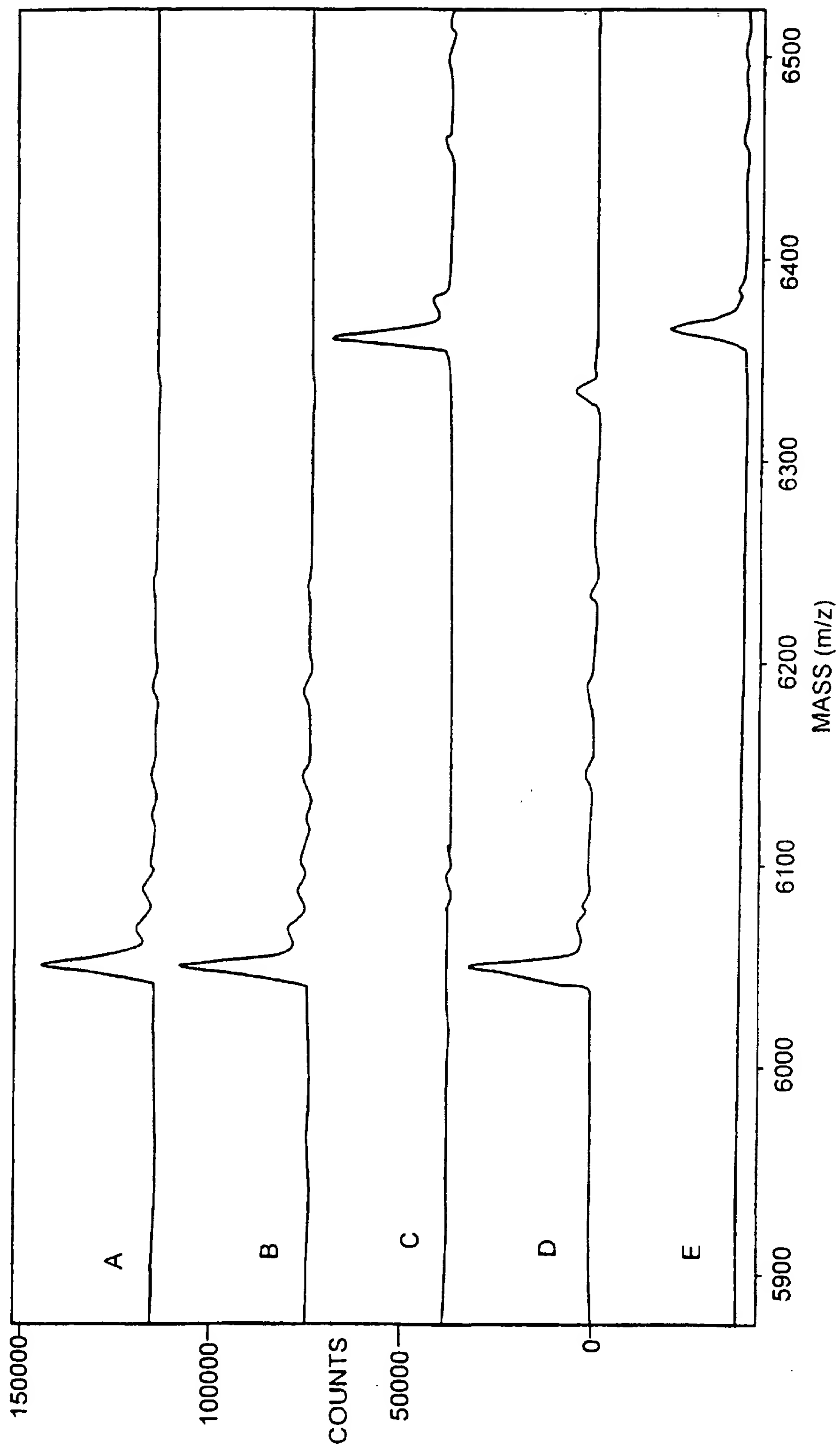
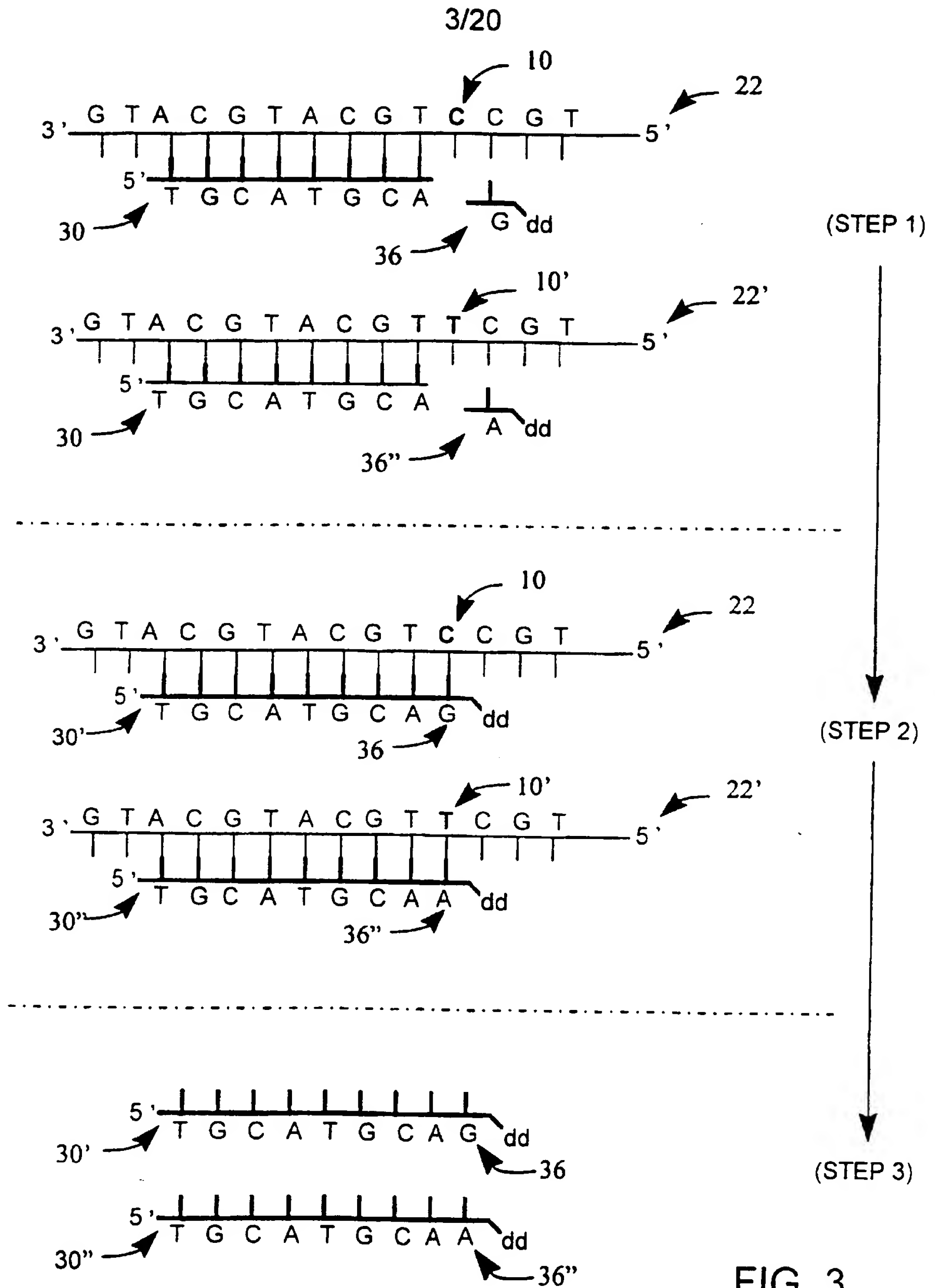


FIG.2



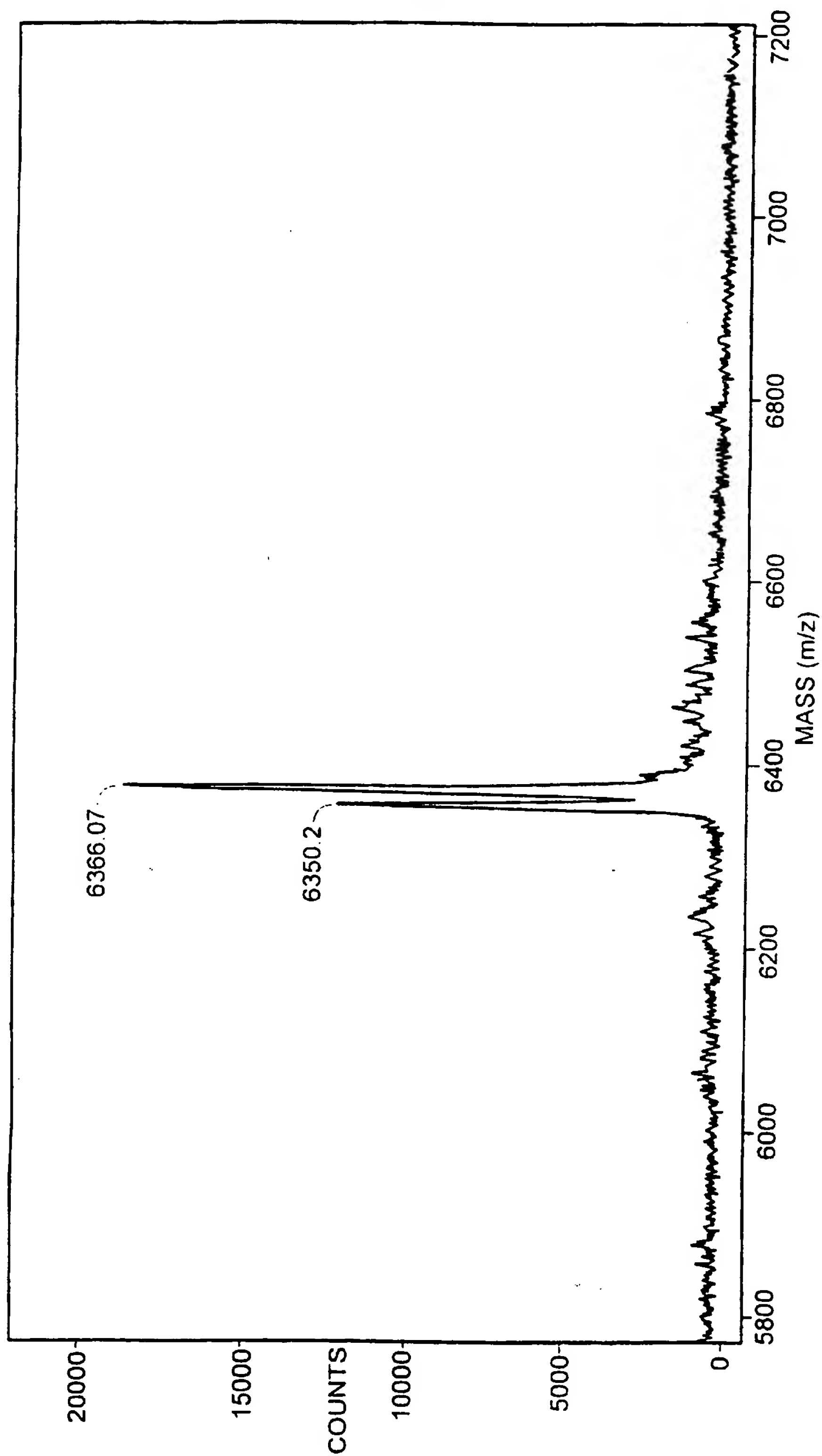


FIG. 4

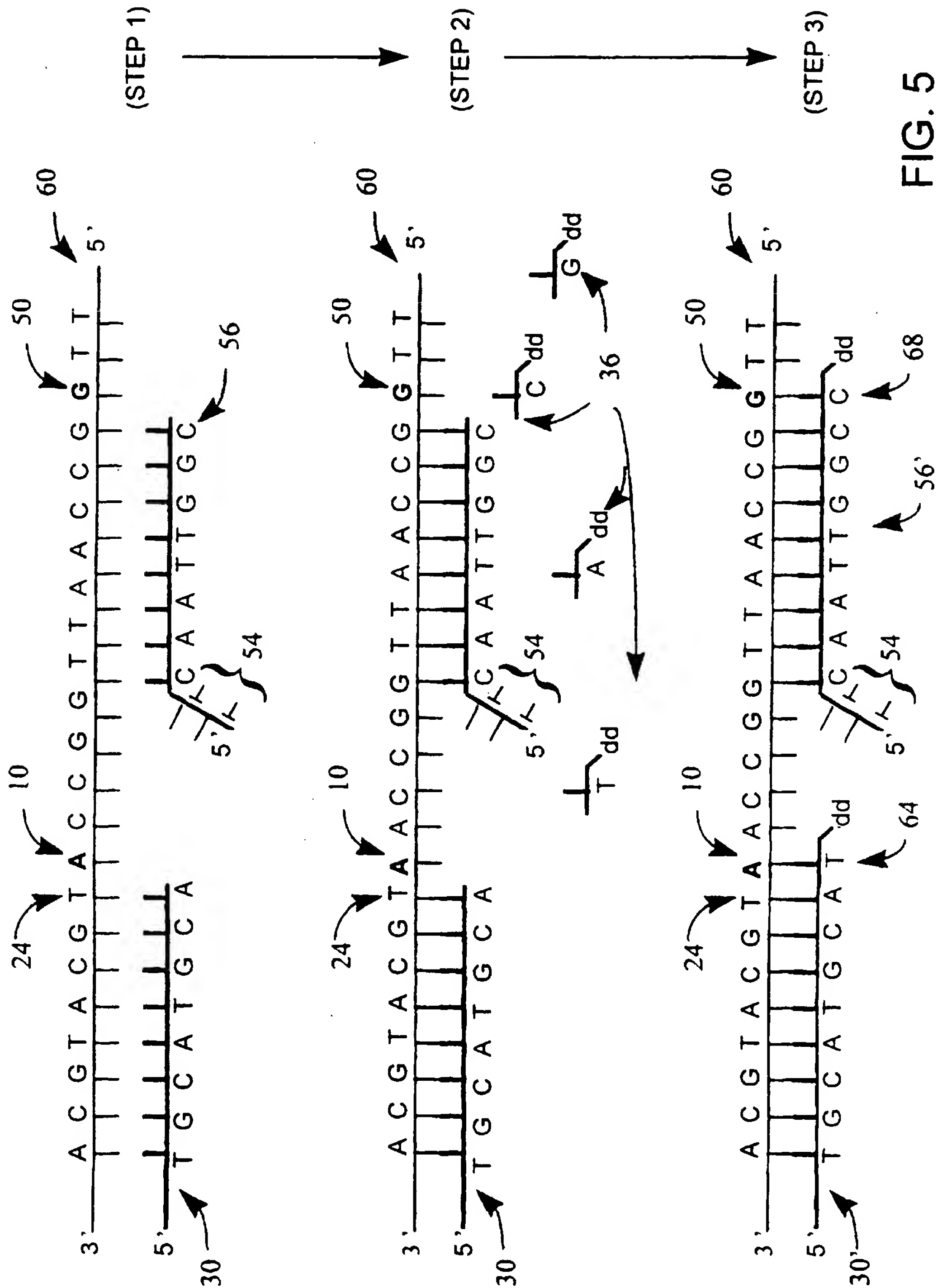


FIG. 5

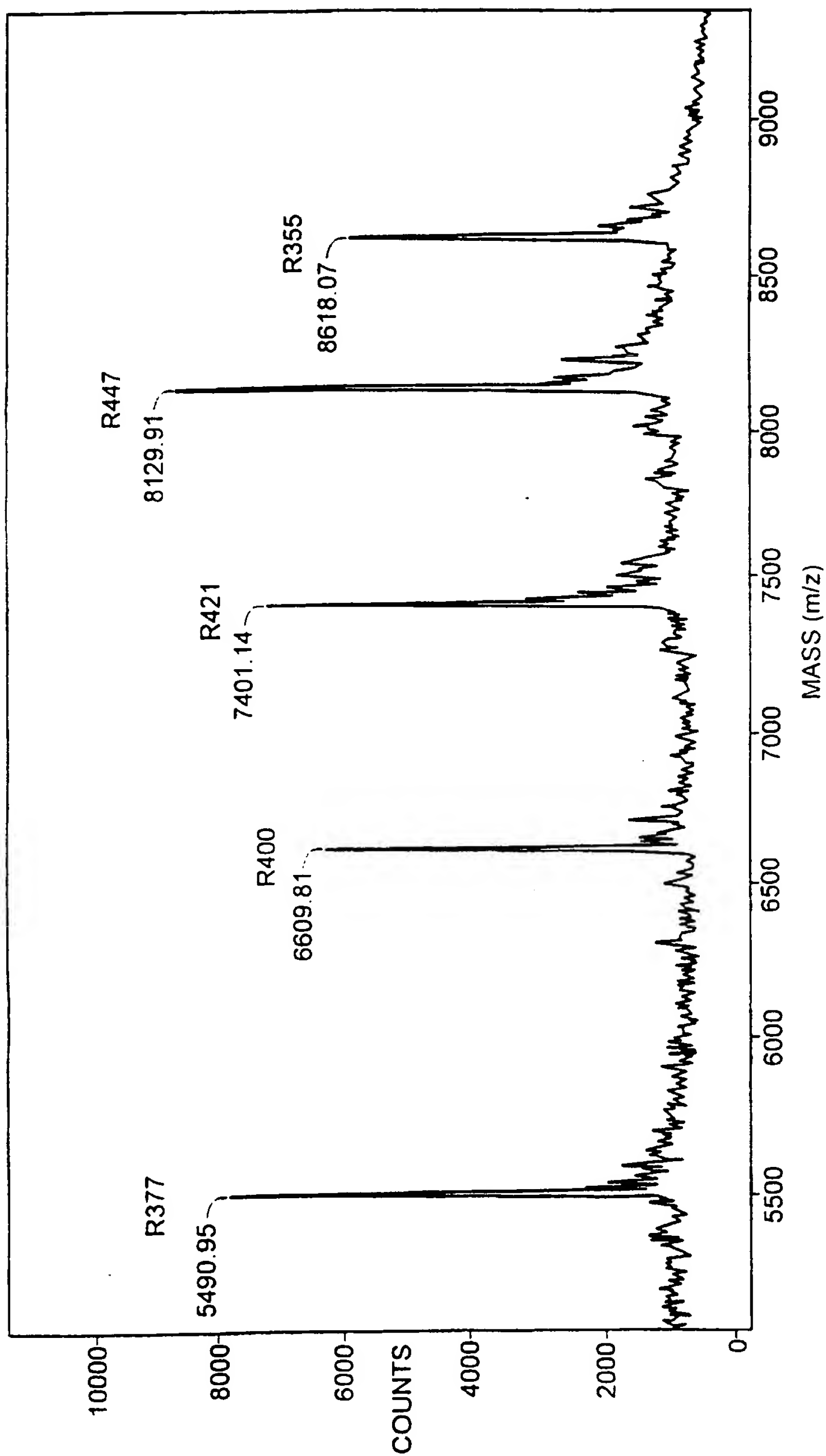


FIG. 6a

7/20

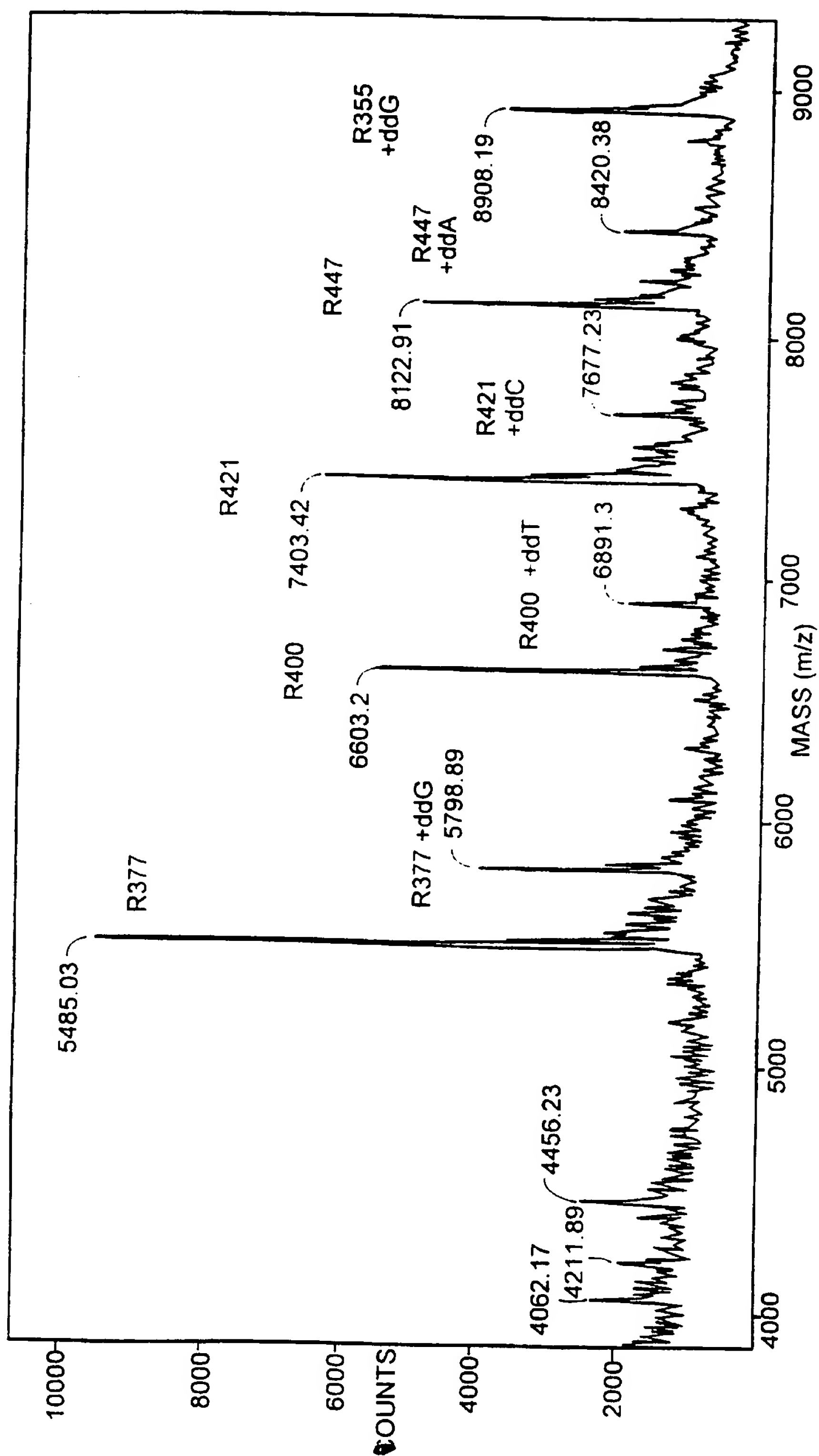


FIG. 6b

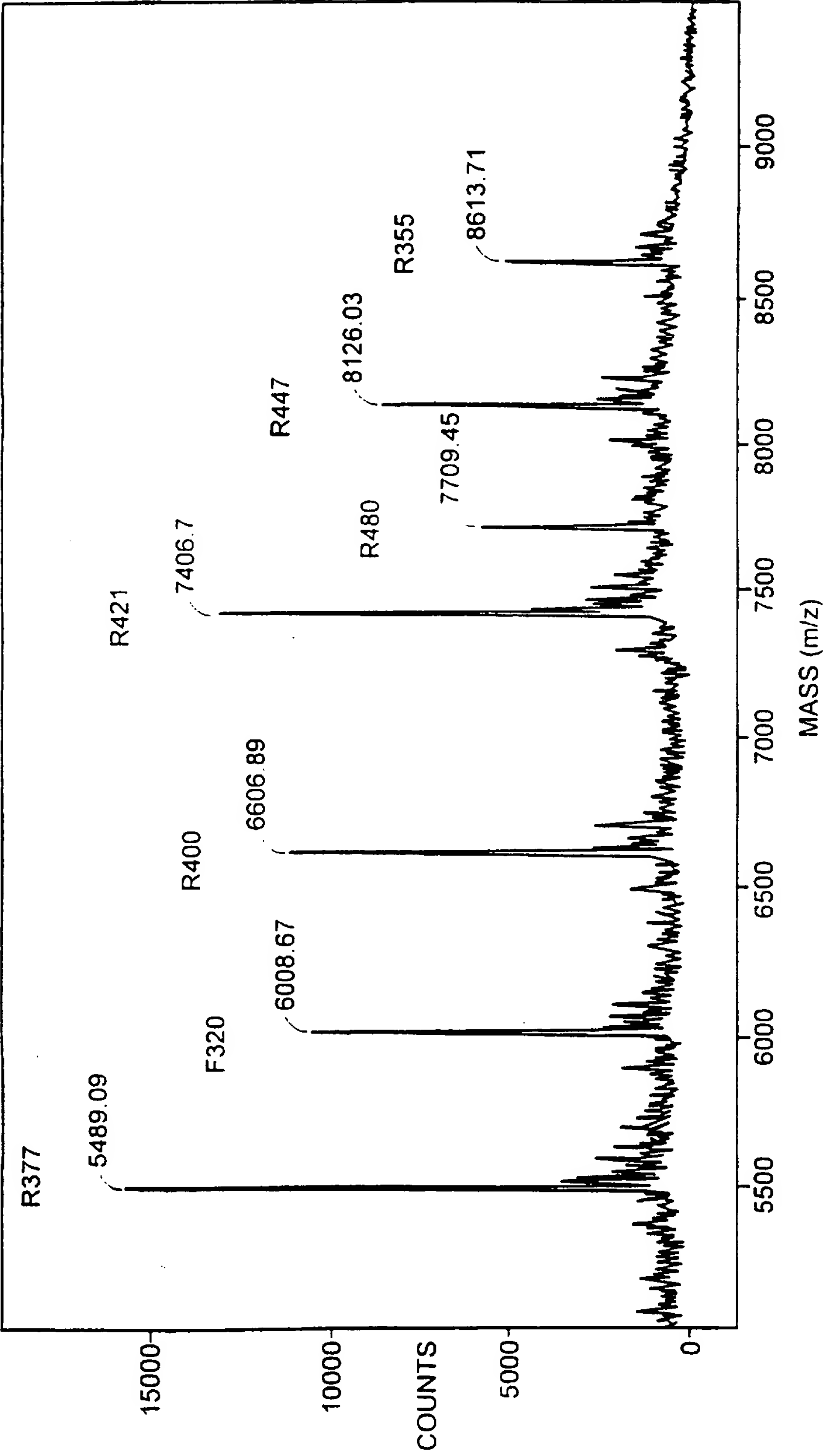


FIG. 6C

9/20

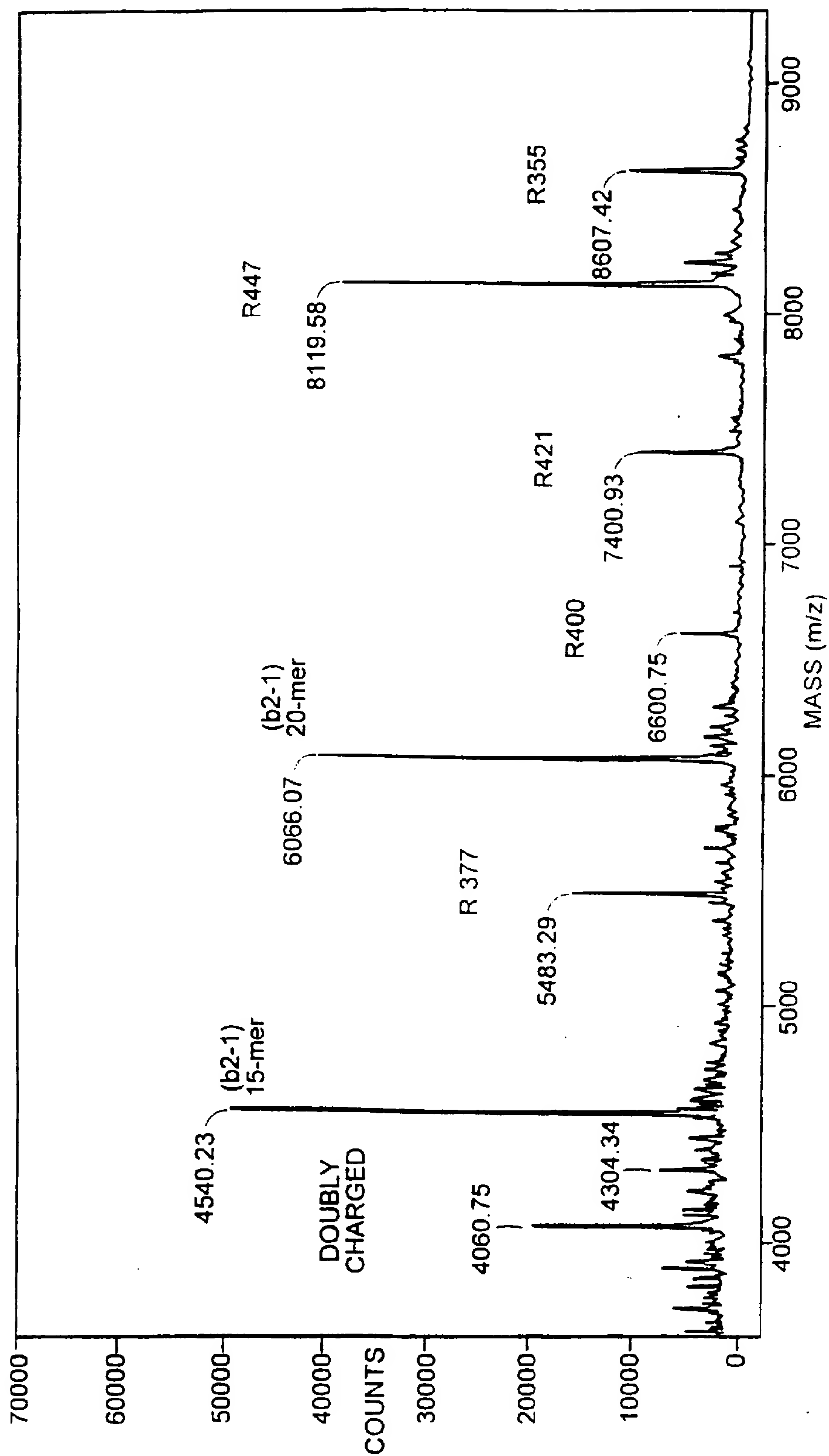


FIG. 6d

10/20

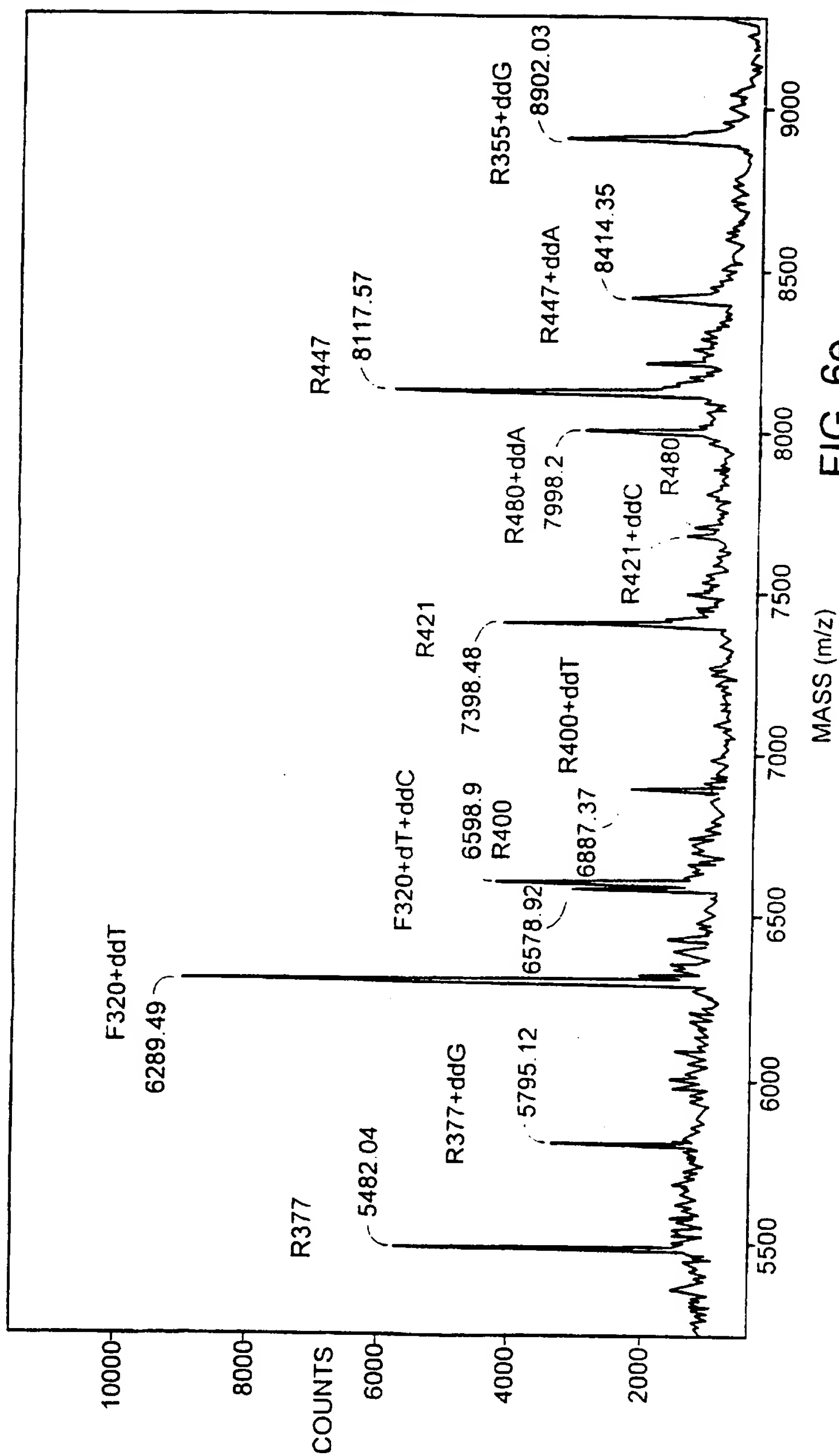


FIG. 6e

11/20

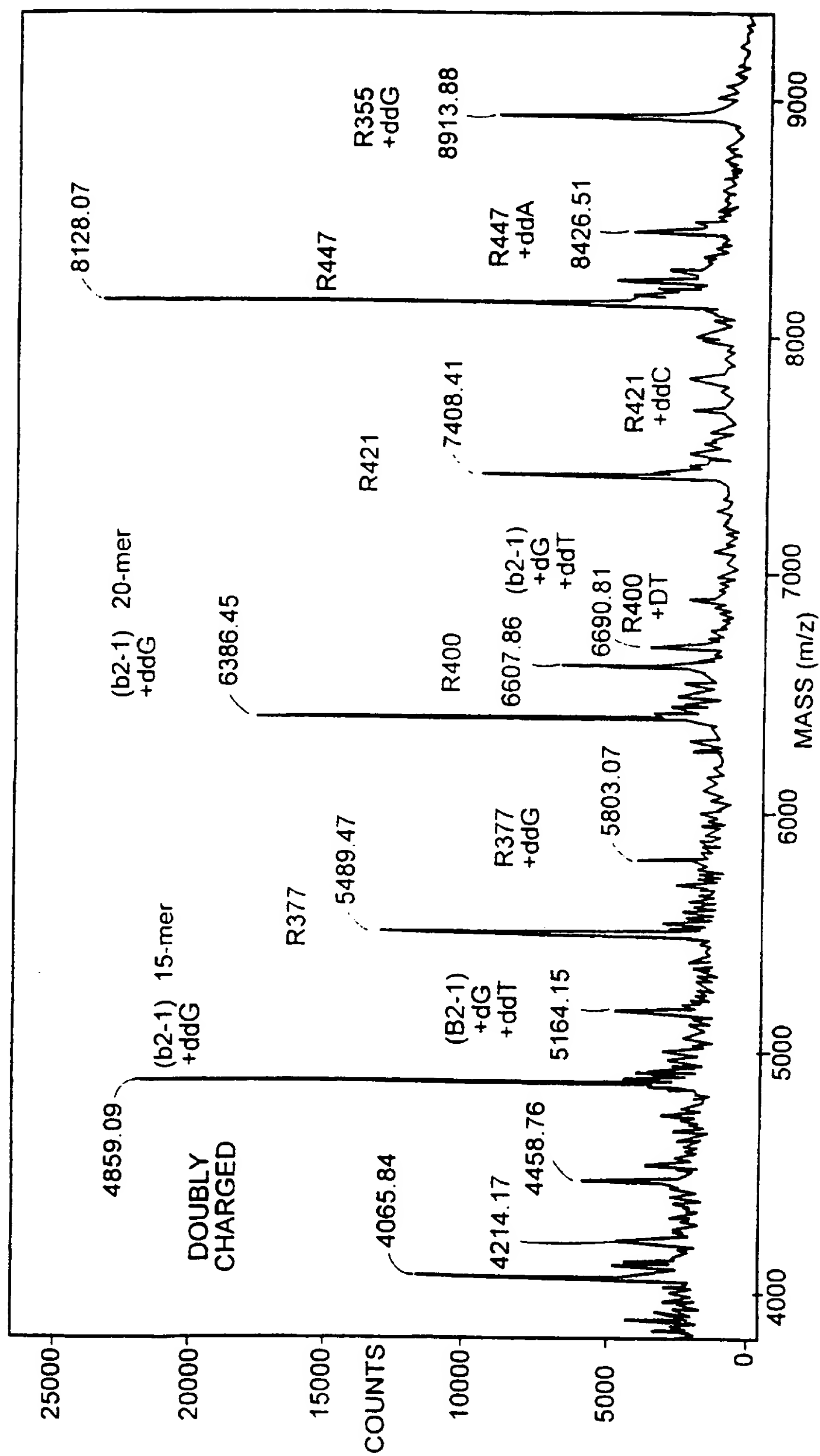


FIG. 6f

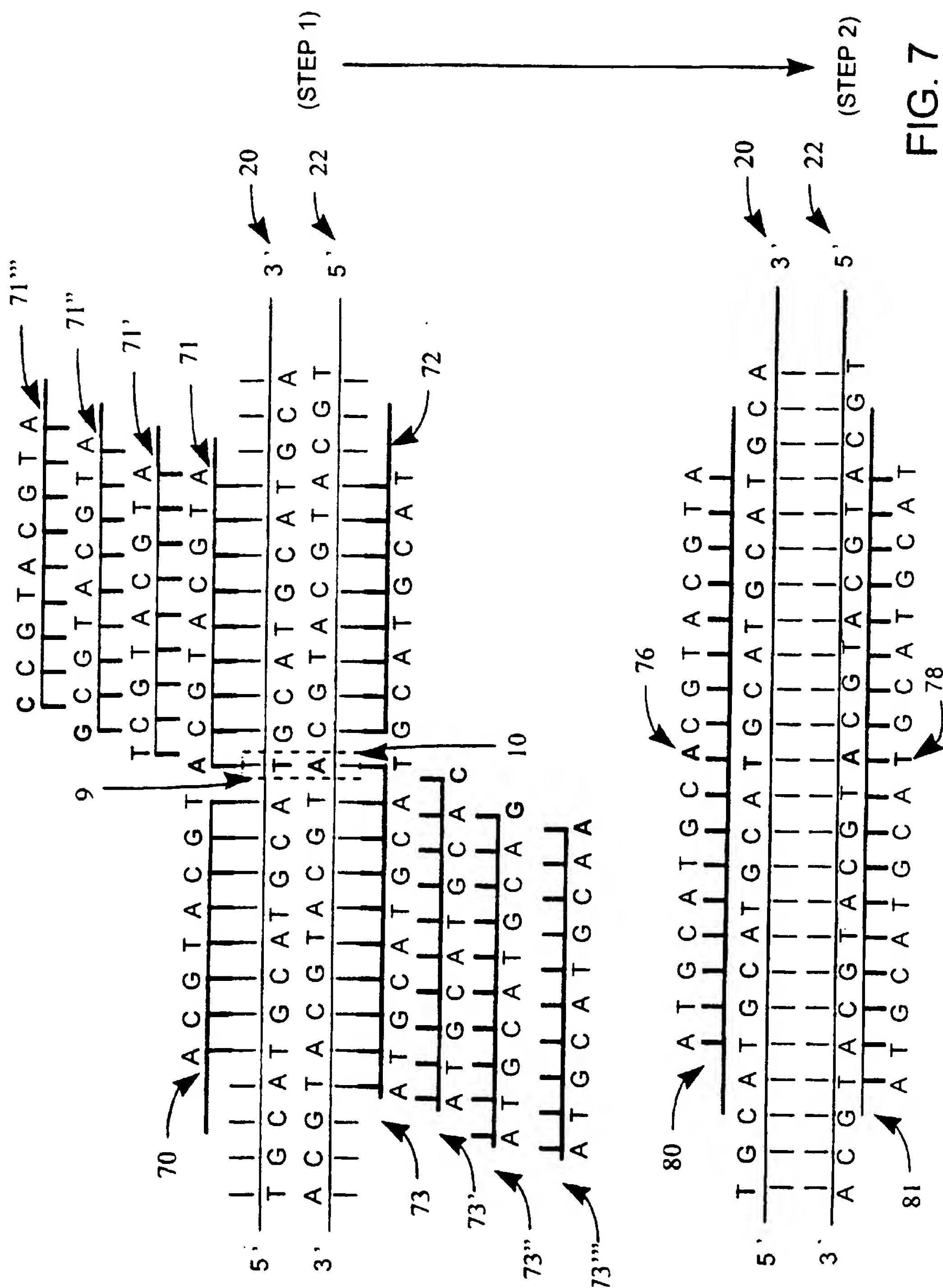


FIG. 7

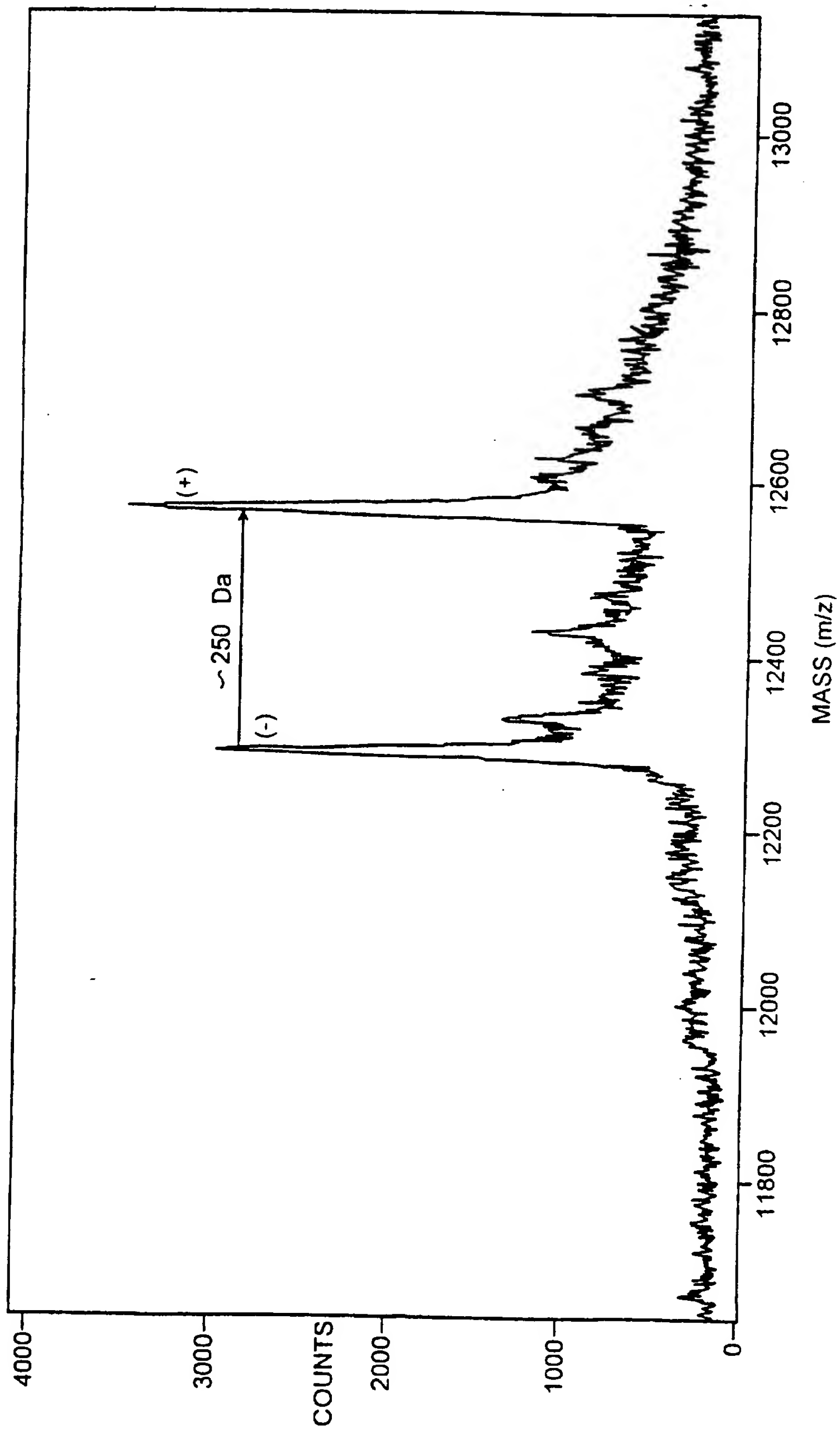


FIG.8

14/20

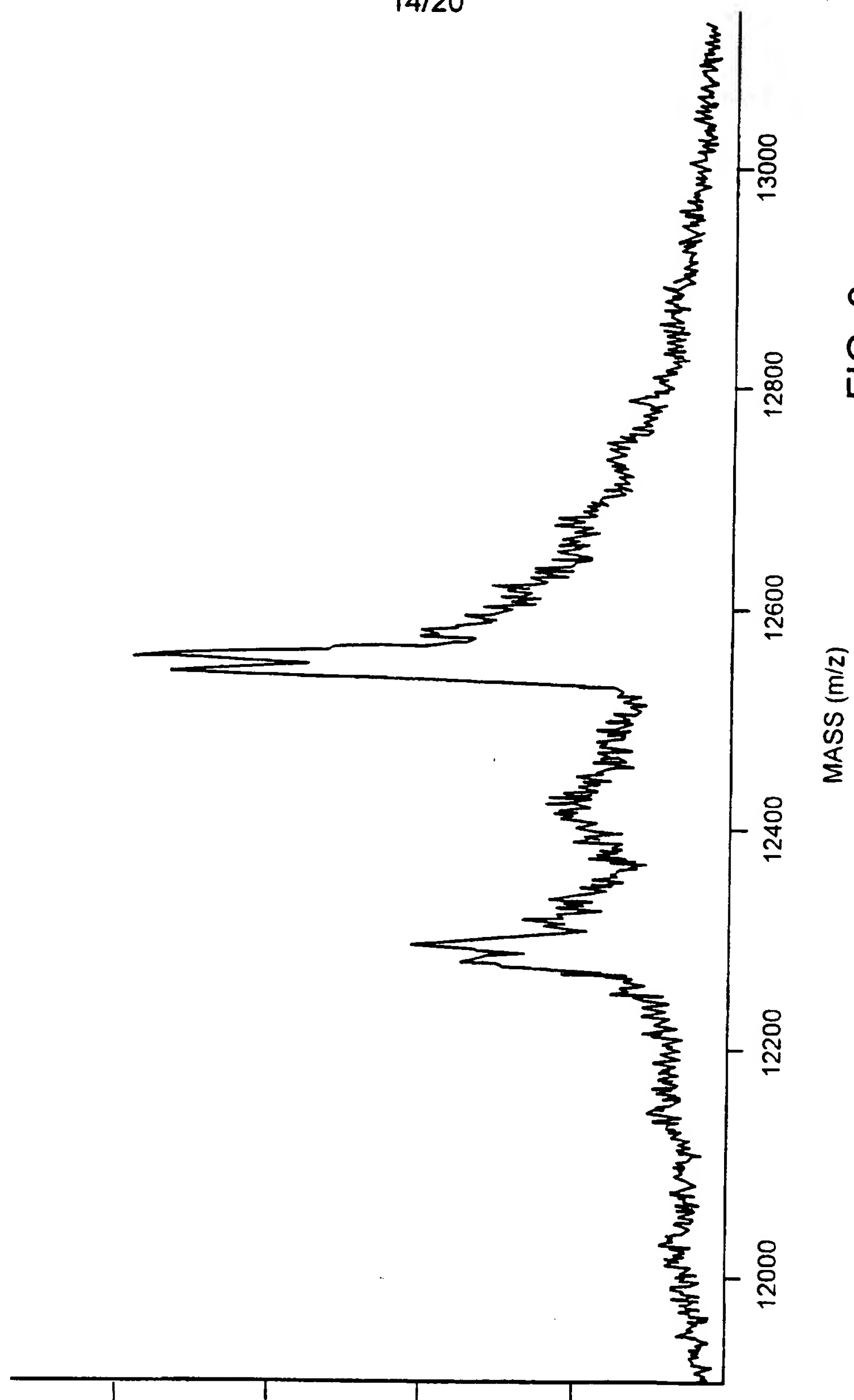
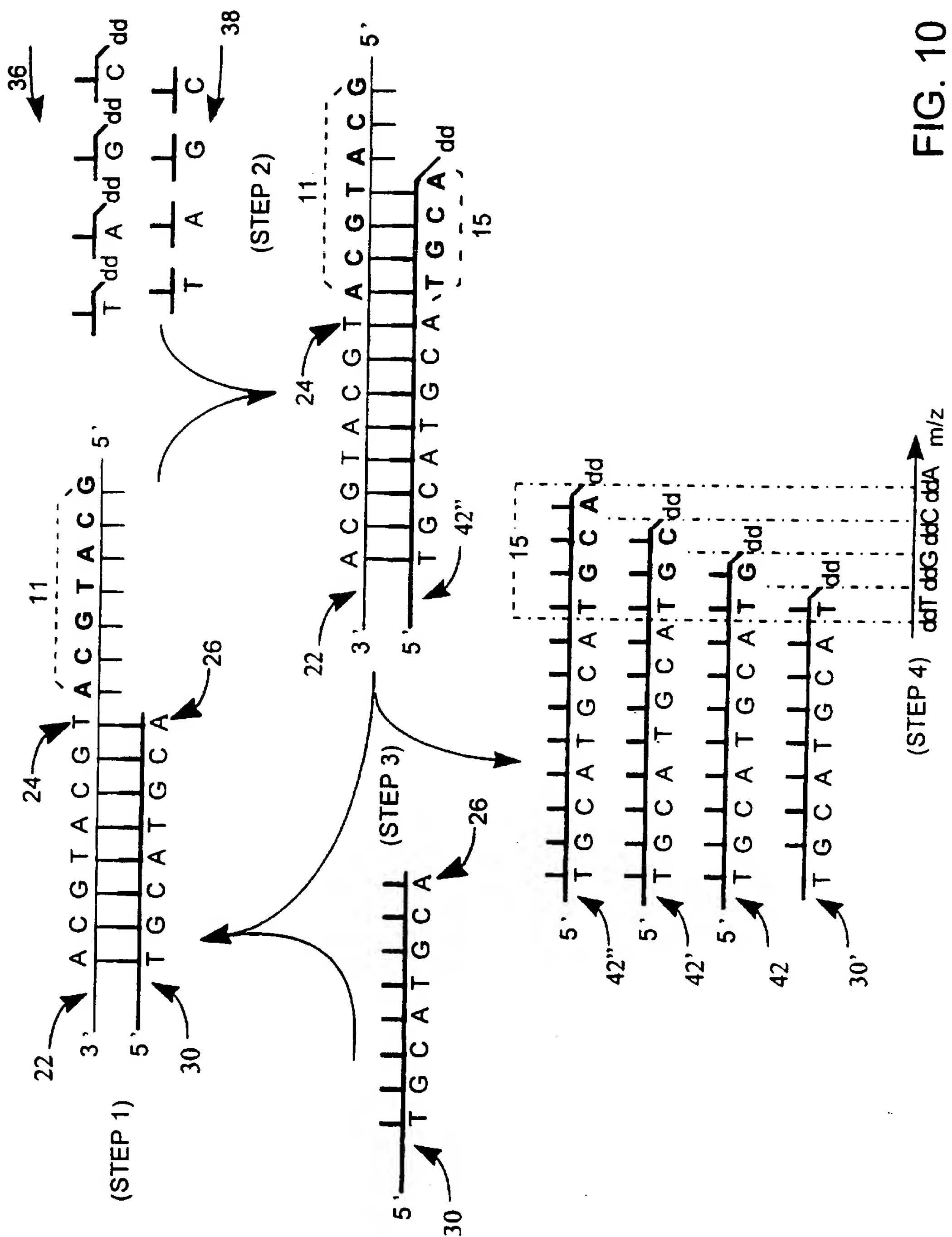


FIG. 9



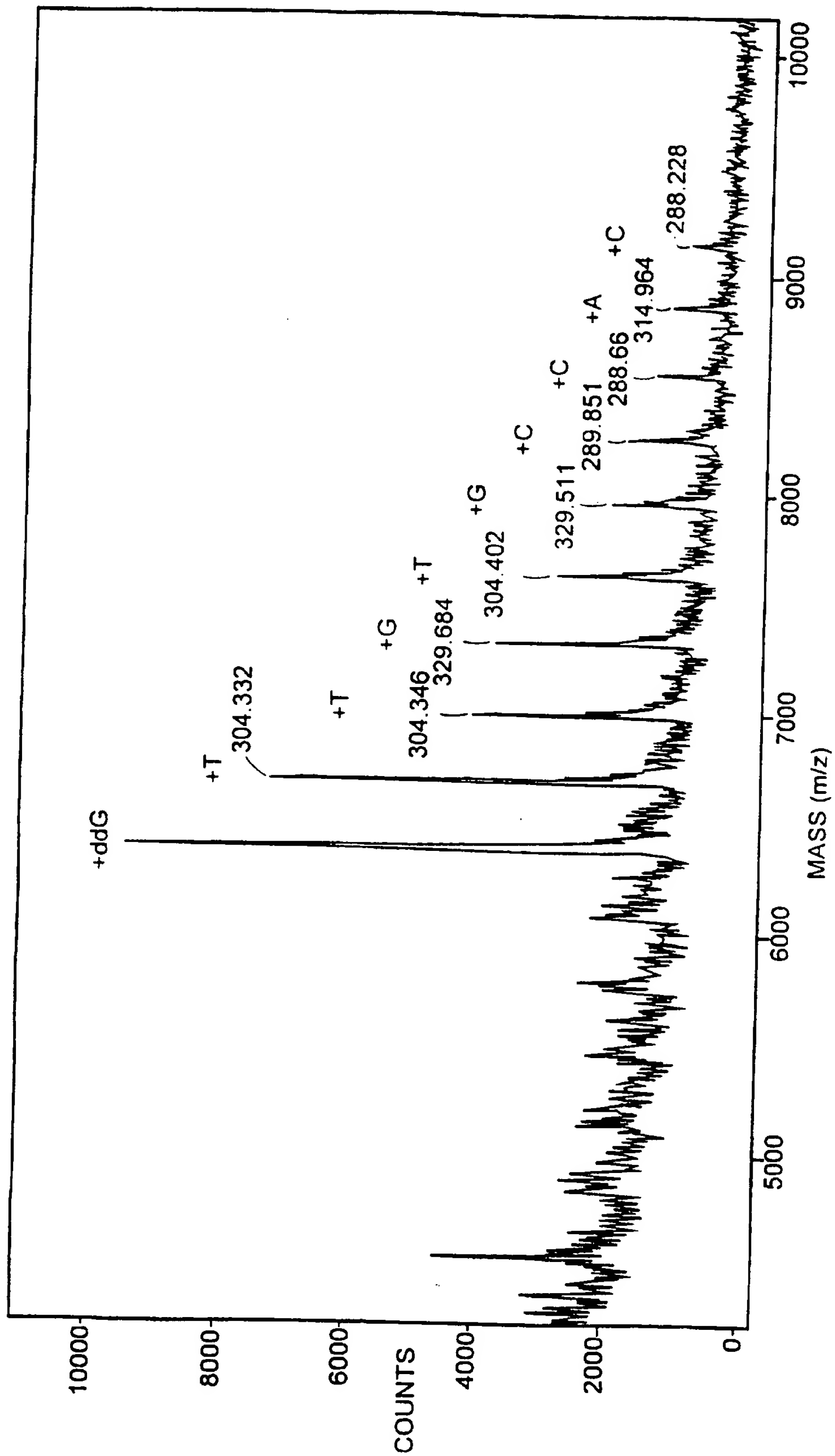


FIG. 11

17/20

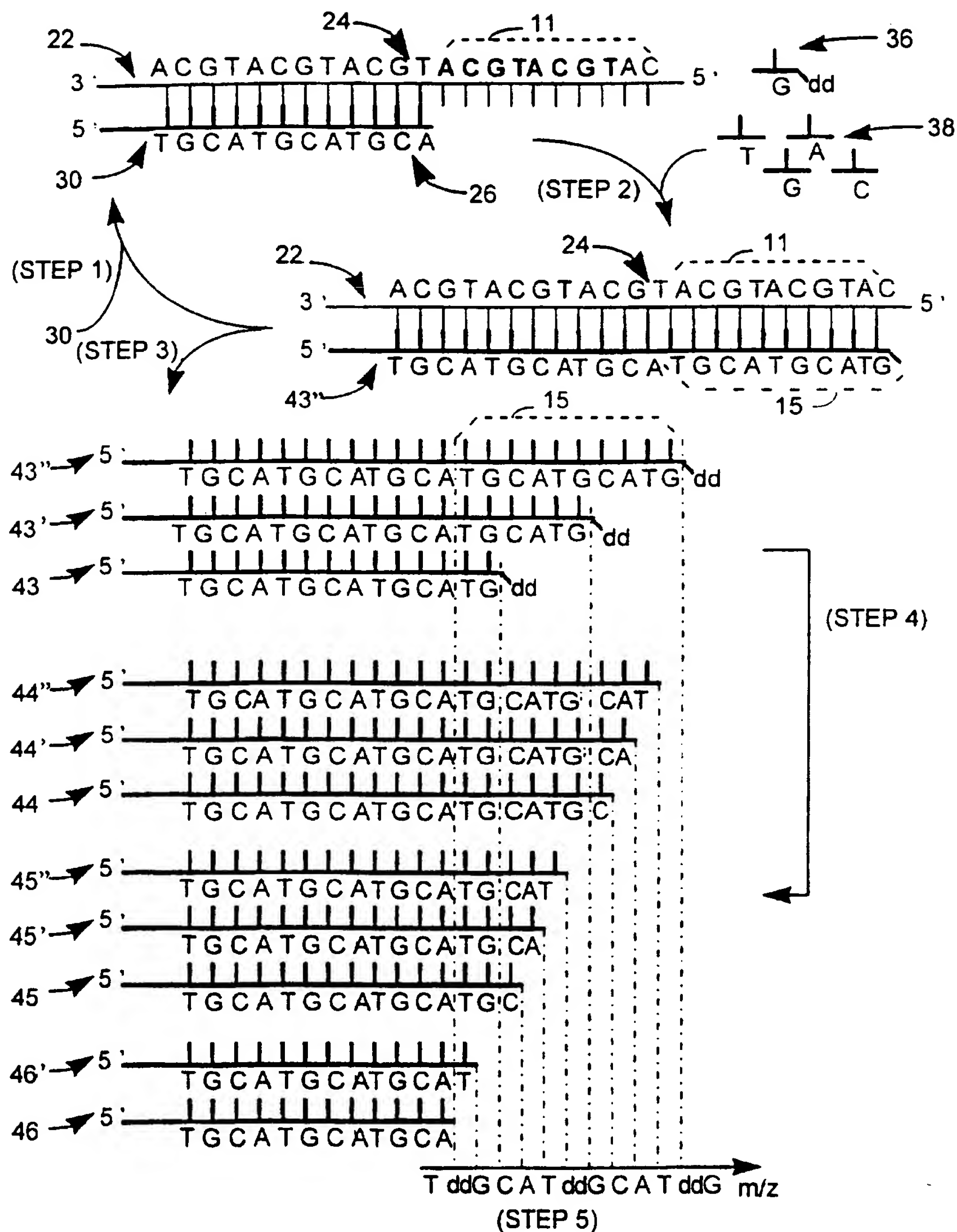
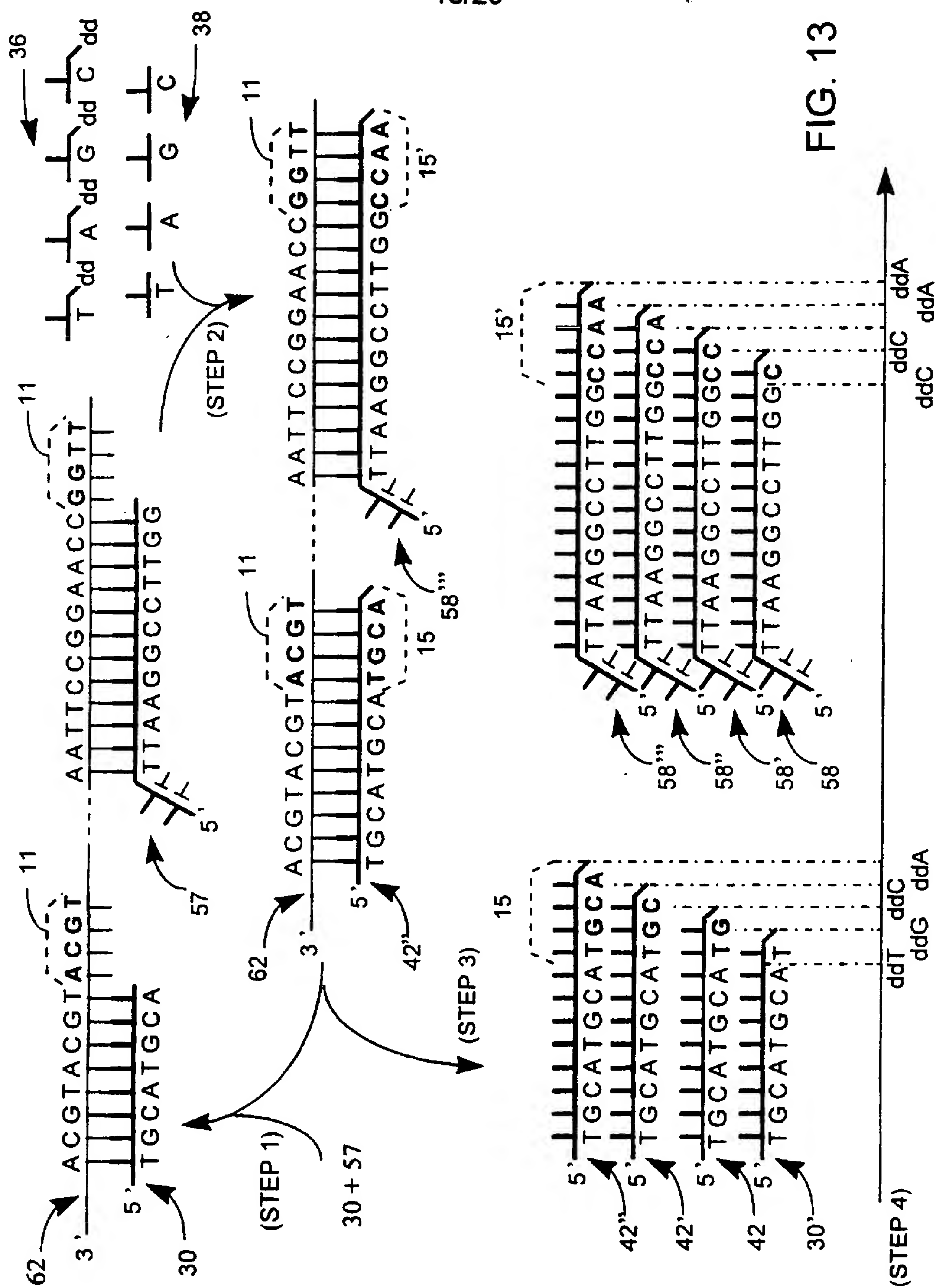


FIG. 12



19/20

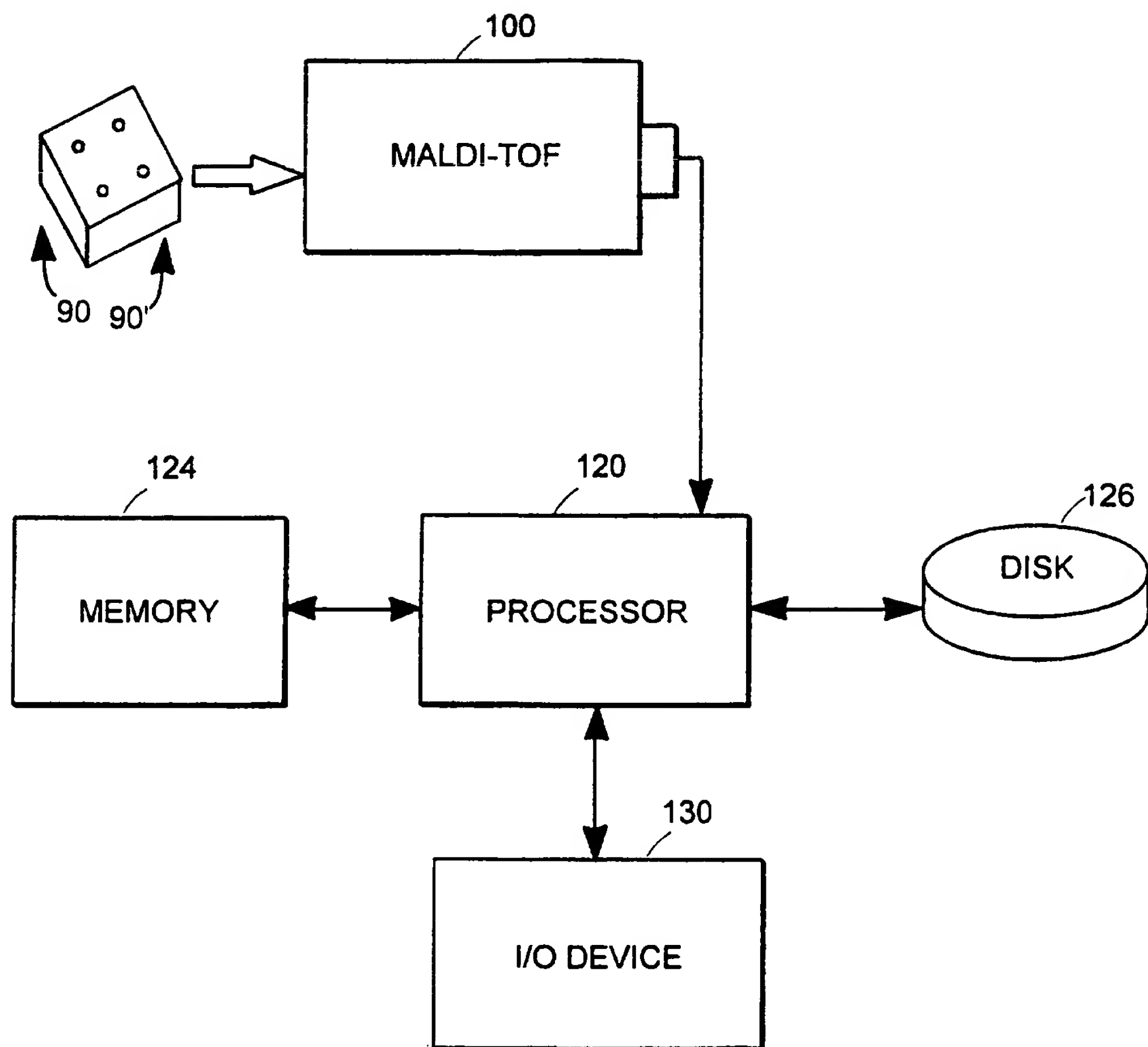


FIG. 14

20/20

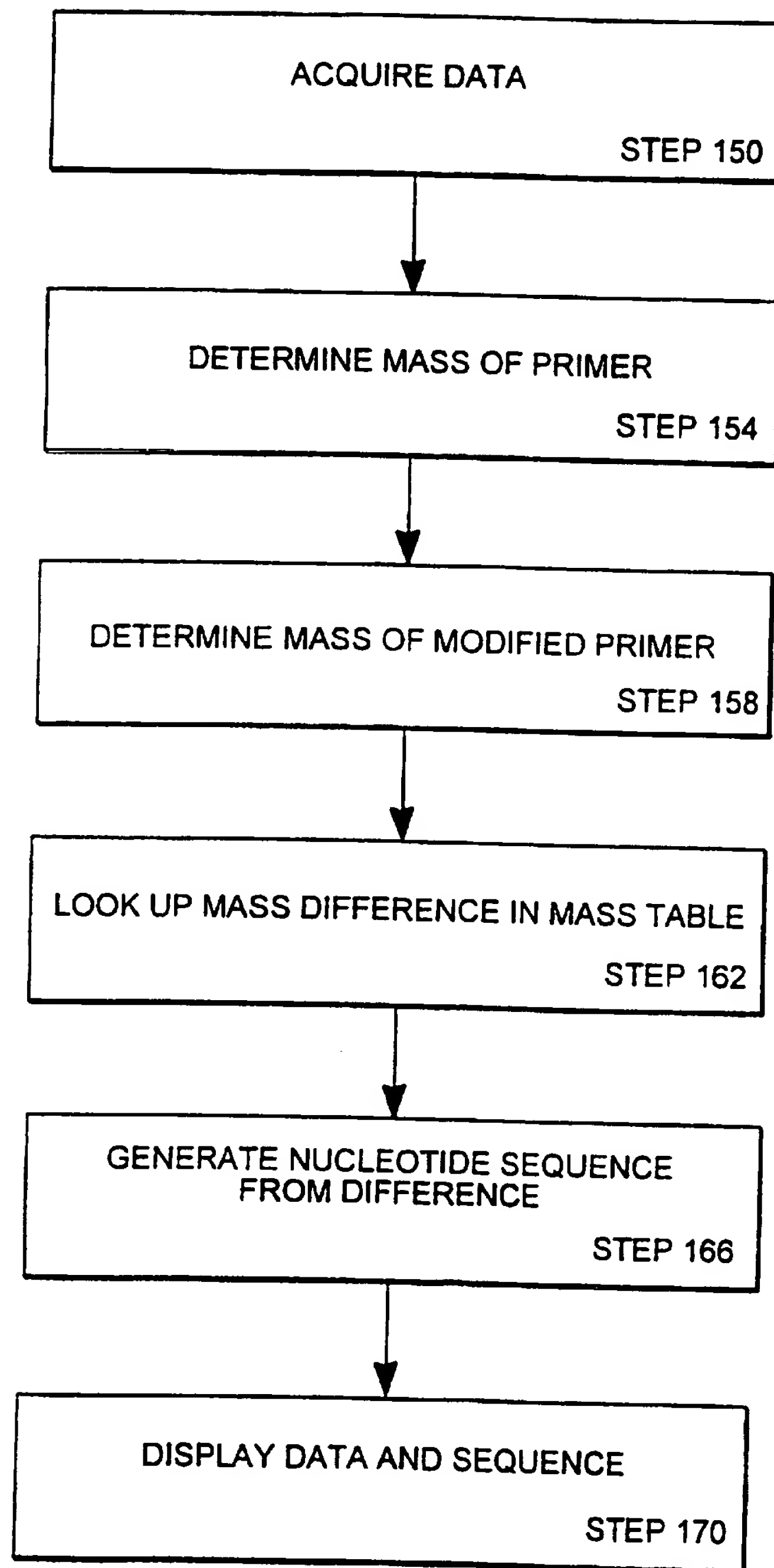


FIG. 15

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/17877

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| X | WO 94 16101 A (KOESTER HUBERT) 21 July 1994 see the whole document --- | 1-11 |
| X | WO 95 14108 A (AMERSHAM INT PLC ;REEVE MICHAEL ALAN (GB); HOWE ROLAND PAUL (GB);) 26 May 1995 see the whole document --- | 1-10 |
| Y | --- | 11 |
| X | WO 96 29431 A (SEQUENOM INC) 26 September 1996 see page 24, line 9 - page 26, line 34 --- | 1-11 |
| X | WO 96 27681 A (IMP CANCER RES TECH ;GUT IVO GLYNNE (GB); BECK STEPHAN AUGUST (GB)) 12 September 1996 see the whole document --- | 1-10 |
| Y | --- | 11 |
| | --- -/-- | |



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

25 February 1998

Date of mailing of the international search report

03/03/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Osborne, H

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/17877

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| Y | SMIRNOV I P ET AL: "SEQUENCING OLIGONUCLEOTIDES BY EXONUCLEASE DIGESTION AND DELAYED EXTRACTION MATRIX-ASSISTED LASER DESORPTION IONIZATION TIME-OF-FLIGHT MASS SPECTROMETRY" ANALYTICAL BIOCHEMISTRY, vol. 238, no. 1, 15 June 1996, pages 19-25, XP000597084 see the whole document | 1-11 |
| Y | WO 96 30545 A (APPLIED GENETICS INC) 3 October 1996 see claims 1-37 | 1-10 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/17877

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|--|--|
| WO 9416101 A | 21-07-94 | AU 5992994 A CA 2153387 A EP 0679196 A JP 8509857 T US 5547835 A US 5605798 A US 5691141 A | 15-08-94 21-07-94 02-11-95 22-10-96 20-08-96 25-02-97 25-11-97 |
| WO 9514108 A | 26-05-95 | EP 0765401 A JP 9505397 T | 02-04-97 27-05-97 |
| WO 9629431 A | 26-09-96 | US 5605798 A AU 5365196 A CA 2214359 A EP 0815261 A | 25-02-97 08-10-96 26-09-96 07-01-98 |
| WO 9627681 A | 12-09-96 | EP 0813609 A | 29-12-97 |
| WO 9630545 A | 03-10-96 | AU 5296496 A | 16-10-96 |